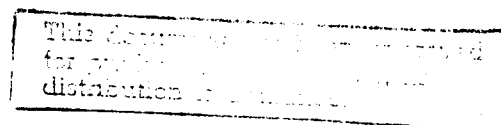


SATELLITE SURFACE MATERIAL COMPOSITION
FROM SYNTHETIC SPECTRA

THESIS
Eugene L. Caudill
Captain, USAF

AFIT/GEO/ENG/94-D02



DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

19941228 023

AFIT/GEO/ENG/94-D02

SATELLITE SURFACE MATERIAL COMPOSITION
FROM SYNTHETIC SPECTRA

THESIS
Eugene L. Caudill
Captain, USAF

AFIT/GEO/ENG/94-D02

Accession For	
NTIS CRAFT	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Avail. To Public	
Dist	Avail. To Public
A-1	

DTIC QUALITY INSPECTED 2

Approved for public release; distribution unlimited

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

AFIT/GEO/ENG/94-D02

SATELLITE SURFACE MATERIAL COMPOSITION
FROM SYNTHETIC SPECTRA

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Master of Science

Eugene L. Caudill, B.S.
Captain, USAF

December, 1994

Approved for public release; distribution unlimited

Acknowledgements

In performing this research and writing this thesis, I received help from several people. I am eternally grateful to my thesis advisor and committee chairman, Major Michael Roggemann. His patience and guidance was instrumental to the successful accomplishment of the research objectives. I am also deeply indebted to the other two members of my thesis committee, Dr Byron Welsh and Dr Steven Rogers. The expertise they exhibit in their fields provided an indispensable source of knowledge. I also wish to thank my family, especially my wife, Joni. Their support and understanding will allow me to look back on my AFIT tour as a very enjoyable experience.

Eugene L. Caudill

Table of Contents

	Page
Acknowledgements	ii
List of Figures	vi
List of Tables	x
Abstract	xi
 I. Introduction	 1
1.1 Motivation	1
1.2 Research Objectives	2
1.3 Assumptions	3
1.4 Summary of Key Results	5
1.5 Organization of Thesis	6
 II. The Sagnac Interferometer	 7
2.1 Introduction to Fourier Transform Spectroscopy	7
2.2 Introduction to the Sagnac Interferometer	13
2.3 SNR of the Sagnac Interferometer	15
2.4 Expected SNRs using Radiometric Models	22
2.5 Summary	24
 III. Theory for Estimation of Percent Composition	 25
3.1 Introduction to the System Architectures	27
3.2 Concepts for Material Classification	28
3.2.1 Feature Extraction and Data Analysis	30

	Page
3.2.2 The Parzen Classifier	31
3.2.3 The Multi-Layer Perceptron	34
3.3 Concepts for Constrained Linear Estimation	39
3.4 Summary	42
IV. The Single Material Problem: Classification	43
4.1 Data Simulation	43
4.2 Data Analysis and Feature Extraction	47
4.2.1 Principal Component Analysis	47
4.2.2 Fisher Discriminants	49
4.3 Performance of the Classifiers	52
4.3.1 Results for the Parzen Classifier	54
4.3.2 Results for the Multi-Layer Perceptron	55
4.3.3 Comparisons of Results for the Two Classifiers	57
4.4 Summary	58
V. The Two Material Problem: Abundance Estimation	59
5.1 The Three Architectures Revisited	59
5.2 MLP Approach	60
5.3 Data Simulation, Analysis, and Feature Extraction	62
5.3.1 Data Simulation	62
5.3.2 Principal Component Analysis	63
5.3.3 Fisher Discriminants	66
5.4 Linear Estimator Approach	68
5.5 Hybrid Approach	69
5.6 Summary	75
VI. Conclusions and Recommendations for Future Research	77

	Page
Appendix A. Analysis and Results	79
A.1 Fisher Discriminants for Single Material Data	80
A.2 Two Material Data shown Projected in Two Principal Component Plane	82
A.3 ANN Rankings for Two Materials whose Spectra are in the Compos- ite Spectrum	95
A.4 Errors in Estimation of Percent Composition: Hybrid Approach . .	98
 Bibliography	 101
 Vita	 103

List of Figures

Figure	Page
1. Michelson Interferometer with Auxiliary Optics	8
2. Interferogram obtained from a monochromatic source	9
3. Interferogram obtained from a polychromatic source	10
4. Schematic Diagram of the Sagnac Interferometer	14
5. Satellite Observation Scenario	18
6. Block Diagram of Three Proposed Systems	27
7. Feature Space for Tank and Jeep Example	29
8. Input/Output Diagram of a Single Perceptron	34
9. Sigmoid and Hyperbolic Tangent Functions	36
10. Interconnection Diagram for a Multi-Layer Perceptron	37
11. Spectral Reflectivities of the Nine Materials	44
12. Single Material Training Data Projected into Plane of First Two Principal Components; Low SNR Case	49
13. Single Material Training Data Projected into Plane of First Two Principal Components; Mid SNR Case	50
14. Single Material Training Data Projected into Plane of First Two Principal Components; High SNR Case	51
15. MLP Architecture for Single Material Problem	56
16. MLP Architecture for Two Material Problem	61
17. Estimation of Percent Composition using the MLP	62
18. Endmembers Projected into PC Plane	64
19. Material 1; Two Material Training Data Projected into PC Plane: Low SNR Case	65
20. Material 1; Two Material Training Data Projected into PC Plane: Mid SNR Case	66

Figure	Page
21. Material 1; Two Material Training Data Projected into PC Plane: High SNR Case	67
22. Block Diagram for Hybrid Approach to Estimation of Percent Composition .	70
23. ANN Rankings for First Material in Two Material Composite for the Low SNR Case	72
24. ANN Rankings for Second Material in Two Material Composite for the Low SNR Case	73
25. Material 2; Two Material Training Data Projected into PC Plane: Low SNR Case	83
26. Material 2; Two Material Training Data Projected into PC Plane: Mid SNR Case	83
27. Material 2; Two Material Training Data Projected into PC Plane: High SNR Case	84
28. Material 3; Two Material Training Data Projected into PC Plane: Low SNR Case	84
29. Material 3; Two Material Training Data Projected into PC Plane: Mid SNR Case	85
30. Material 3; Two Material Training Data Projected into PC Plane: High SNR Case	85
31. Material 4; Two Material Training Data Projected into PC Plane: Low SNR Case	86
32. Material 4; Two Material Training Data Projected into PC Plane: Mid SNR Case	86
33. Material 4; Two Material Training Data Projected into PC Plane: High SNR Case	87
34. Material 5; Two Material Training Data Projected into PC Plane: Low SNR Case	87
35. Material 5; Two Material Training Data Projected into PC Plane: Mid SNR Case	88
36. Material 5; Two Material Training Data Projected into PC Plane: High SNR Case	88

Figure		Page
37.	Material 6; Two Material Training Data Projected into PC Plane: Low SNR Case	89
38.	Material 6; Two Material Training Data Projected into PC Plane: Mid SNR Case	89
39.	Material 6; Two Material Training Data Projected into PC Plane: High SNR Case	90
40.	Material 7; Two Material Training Data Projected into PC Plane: Low SNR Case	90
41.	Material 7; Two Material Training Data Projected into PC Plane: Mid SNR Case	91
42.	Material 7; Two Material Training Data Projected into PC Plane: High SNR Case	91
43.	Material 8; Two Material Training Data Projected into PC Plane: Low SNR Case	92
44.	Material 8; Two Material Training Data Projected into PC Plane: Mid SNR Case	92
45.	Material 8; Two Material Training Data Projected into PC Plane: High SNR Case	93
46.	Material 9; Two Material Training Data Projected into PC Plane: Low SNR Case	93
47.	Material 9; Two Material Training Data Projected into PC Plane: Mid SNR Case	94
48.	Material 9; Two Material Training Data Projected into PC Plane: High SNR Case	94
49.	ANN Rankings for First Material in Two Material Composite for the Mid SNR Case	96
50.	ANN Rankings for Second Material in Two Material Composite for the Mid SNR Case	96
51.	ANN Rankings for First Material in Two Material Composite for the High SNR Case	97

Figure		Page
52.	ANN Rankings for Second Material in Two Material Composite for the High SNR Case	97

List of Tables

Table		Page
1.	Values for well-known constants	19
2.	Values chosen for radiometric models	22
3.	Radiometric scenarios and results	23
4.	Average SNRs for Single Material Data	46
5.	Principal Eigenvalues for Material Database	48
6.	Fisher Discriminants for Single Material Data; Low SNR Case	53
7.	Confusion Matrix for Single Material Problem obtained from Parzen Classifier; Low SNR Case	55
8.	Confusion Matrix for Single Material MLP Classifier	57
9.	Classification Accuracies for Single Material Problem	57
10.	Generalized Fisher Discriminants for Two Material Problem	68
11.	Errors in Estimates using Linear Technique	69
12.	Error Rates for Confusion Matrix Resulting from MLP Classifier	71
13.	Errors in Estimates using Hybrid Approach; Low SNR Case	74
14.	Comparison of mean RMS Errors; Linear Approach vs Hybrid Approach . . .	76
15.	Fisher Discriminants for Single Material Data; Mid SNR Case	80
16.	Fisher Discriminants for Single Material Data; High SNR Case	81
17.	Errors in Estimates using Hybrid Approach; Mid SNR Case	99
18.	Errors in Estimates using Hybrid Approach; High SNR Case	100

Abstract

The objective of this research was to determine if measurements from a Sagnac interferometer could provide reliable estimates of satellite material composition. The Sagnac interferometer yields a spatial interferogram that can be sampled by a linear detector array. The interferogram is related to the spectrum of the source through a Fourier transform. Here, spectral reflectivities of nine common satellite materials were used to simulate the spectrum one obtains from an ideal Sagnac interferometer in the beam-train of a ground-based telescope whose mission is to view satellites. The signal-to-noise ratio of the spectrum was varied to simulate the effect of range variation between the sensor and the satellite. The simulated spectra consisted of a linear mixture of spectra from two of the nine materials.

Three different architectures were developed and their performances compared. One of the three architectures consisted of nine artificial neural networks (ANNs), one for each material, and a linear estimator that estimated the satellite surface area attributable to each material. This method estimates the material composition by using a classifier to identify the materials contributing to the mixture, then eliminating unlikely contributors to the mixture before performing a constrained linear estimate. It is shown that due to high classification errors, the system using solely a linear estimator provides the estimate with the lowest errors.

SATELLITE SURFACE MATERIAL COMPOSITION FROM SYNTHETIC SPECTRA

I. Introduction

1.1 Motivation

The Persian Gulf War with Iraq demonstrated the new role that spacecraft will have in all future international conflicts. The missions of these satellites included such things as photoreconnaissance of weather and ground movements, global positioning, and real-time communications. Due to the increased use of space assets in wartime, the United States Space Command must identify each operable space asset as belonging to either a friend or a foe, and assess its mission and health status. An economical means to identify these objects is to use ground-based observation platforms. These systems primarily use two types of measurements, image and radar, for object identification. Both image and radar technologies have made significant improvements in the last few decades, but these technologies are limited in their information content. The field of hyperspectrometry, the simultaneous measurement of both spatial (image) and spectral (wavelength) information, provides additional capabilities and appears useful as another technique for Space Object Identification (SOI). The Phillips Laboratory at Kirtland AFB, New Mexico, is actively pursuing the use of a

hyperspectrometer based on Sagnac interferometer for ground-based observations of orbiting satellites, and is the sponsor of this research.

1.2 Research Objectives

The overall objective of this research is to determine whether the interferogram recorded by a Sagnac triangular-path interferometer may be used to provide reliable estimates of a spacecraft's material composition. Such a determination must be based on the expected signal-to-noise ratio (SNR) of data obtained from the Sagnac interferometer and on the ability to decompose a composite spectrum into its constituent spectra. To meet this objective, several tasks are required:

- Develop a SNR expression for the Sagnac interferometer.
- Model typical satellite observation scenarios and estimate expected range of SNRs.
- Simulate observation data using appropriate SNR levels.
- Develop a system to estimate the spacecraft's material composition.
 - Baseline MLP performance to performance of a Parzen classifier.
 - Design MLP for direct estimate of material composition.
 - Design constrained linear estimator.
 - Design hybrid estimator using a MLP classifier and a constrained linear estimator.
- Compare and evaluate the results.

The first task requires an understanding of the principles of Fourier Transform Spectroscopy (FTS). The development of the SNR expression for the Sagnac interferometer is based on assumptions and conclusions generally accepted for FTS. Once an expression for the SNR is developed, the second task provides realistic SNRs for data obtained from a ground-based Sagnac interferometer whose purpose is to view satellites. The third task uses a material database, containing spectral reflectivities of common satellite materials, to create simulated data with the appropriate SNRs. The data is entered into a system that is designed to estimate the percentage of the satellite's surface that is covered by a material. In this research, three such systems were developed. Each system uses a database of spectral reflectivities of known satellite surface materials. The final task compares and evaluates the results obtained from the three systems.

1.3 Assumptions

In order to meet the objectives of the research in the allotted time, and to find an upper bound on system performance, several assumptions are necessary. Some of the assumptions are the result of engineering judgement, while others are found in the literature. Many of the assumptions are made in order to keep the modelling process simple and manageable.

The first assumption is that the Sagnac interferometer is in the beam-path of a ground-based telescope, whose application is to view satellites, and that the reflected light from the satellite is present on the input aperture of the Sagnac interferometer throughout the observation interval. Sunlight reflected from the surface of the satellite is assumed to be

the sole source of irradiance entering the Sagnac interferometer. Each detector observes the light reflected from more than a single satellite surface resulting in a phenomena known as "spectral mixing" (10:327). Spectral mixing can be of two types, either linear or nonlinear, depending upon the geometry of photon interaction. Linear mixing occurs if a single photon encounters only one material, whereas nonlinear mixing occurs if a single photon is scattered by two or more materials prior to being detected by the sensor (3:2069). Here, it is assumed that spectral mixing occurs in a purely linear fashion.

Light from other sources such as the background and the atmosphere are assumed negligible. The assumption of negligible background light is restrictive, however the effects of this assumption are lessened when satellites are viewed during what is referred to as *terminator*, a one-to-two hour period prior to sunrise or following sunset. During terminator, the sun illuminates the satellite but does not illuminate the sky or the ground-based sensor. Therefore, the sky background can be modelled as a four Kelvin blackbody radiator. A four Kelvin blackbody radiates light of much weaker irradiance than the light reflected off of an illuminated satellite. The other assumption, of negligible atmospheric degradation, is more restrictive because the atmosphere absorbs light energy across the spectrum. However, using excellent atmospheric models and calibration procedures commonly used in FTS, this restriction poses little concern for a properly designed system.

Several additional assumptions are made regarding the equipment design: The Sagnac interferometer is assumed flawless, its optics are free of aberrations and its alignments and calibrations are exact. The output of the interferometer is measured with a charge-coupled

device (CCD) detector array. The individual detector elements in the array are assumed equally sensitive to all wavelengths in the range 300 nm to 1 μm , and entirely insensitive to all other wavelengths. It is further assumed that the post-processing algorithms used on the raw data eliminate all remaining undesirable effects and yield spectra that is exactly that of the satellite plus additive Gaussian noise of appropriate strength.

Other assumptions are necessary and are discussed in the text. These additional assumptions are only applicable to specific segments of the research, and are mainly used to simplify analysis.

1.4 Summary of Key Results

This thesis shows that, under ideal conditions, the interferogram obtained with the Sagnac interferometer may be used in a system that is designed to estimate the material composition of an orbiting spacecraft. Of the systems considered in this research, a neural network, a constrained linear estimator, and a hybrid system with a neural network and a constrained linear estimator, the constrained linear estimator provides the percent composition estimate with the lowest RMS error. On data containing the spectra of two materials, the average error in the estimate for the constrained linear estimator was only 5.55% for data with average signal-to-noise ratios (SNRs) of 10, and 0.68% for data with average SNRs of 100.

1.5 Organization of Thesis

The research objectives are treated in this thesis as follows: Chapter II briefly presents the theory of Fourier Transform Spectroscopy and then the first-order operation of the Sagnac interferometer. Chapter II also presents the derivation of a SNR expression for the Sagnac interferometer, and a model that provides expected SNRs for typical satellite observation scenarios. Chapter III provides block diagrams for three systems that estimate the satellite's surface material composition and then proceeds to discuss the theoretical concepts necessary for their implementation. Areas included in this theoretical discussion are the Multi-Layer Perceptron (MLP), the Parzen classifier, and a constrained linear estimator. Chapter IV provides validation of the MLP for a one-material problem, and Chapter V gives results of applying the three systems to a two-material problem. Chapter VI draws conclusions and provides recommendations for future research.

II. The Sagnac Interferometer

As previously mentioned, the Sagnac interferometer is under investigation as a potential instrument for purposes of Space Object Identification. A fundamental question must be answered, "Can the Sagnac interferometer provide information not currently available in either image or radar measurements?" This research provides an answer by examining the spectra obtained from interferometric measurements using a Sagnac interferometer and determining whether the spectra is of sufficient SNR to be useful for estimates of material composition.

The groundwork for this research effort is presented in this chapter through an estimate of the Sagnac interferometer's SNR for typical ground-based satellite observation scenarios. The sections in this chapter include:

- Introduction to Fourier Transform Spectroscopy
- Introduction to the Sagnac Interferometer
- SNR of the Sagnac Interferometer
- Expected SNRs using Radiometric Models

2.1 Introduction to Fourier Transform Spectroscopy

Fourier Transform Spectroscopy (FTS) is a method to recover the electro-magnetic spectrum of the light emitted or reflected from an object by performing a Fourier transform on the measured interference pattern, or interferogram (27:19). The interferogram is measured

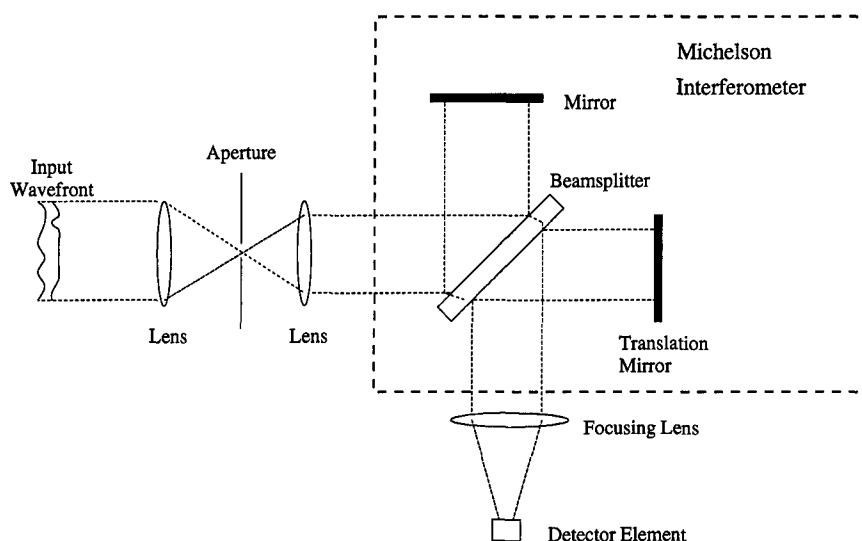


Figure 1. Michelson Interferometer with Auxiliary Optics.

by an interferometer. The interferometer most often used today in FTS is the Michelson interferometer (9:1023A). The Michelson interferometer, shown in Fig. 1, was introduced in 1880 when Dr Albert A. Michelson measured the speed of light in vacuum in the well-known Michelson-Morley experiment (2:17). In the Michelson interferometer, light from the source is focused onto a circular aperture and then collimated. The collimated beam impinges upon the beam-splitter where 50% of the beam is transmitted and 50% of the beam is reflected. The two beams in the two arms of the interferometer are reflected by the mirrors, recombined by the beam-splitter, and then focused onto a single detector (9:1023A).

Consider the condition when light from a monochromatic source is incident on the input aperture. If the beams in the two arms of the interferometer have equal optical path differences (OPD) , they will constructively interfere and the light on the detector will appear bright. As one mirror is translated and the OPDs become unequal, destructive interference

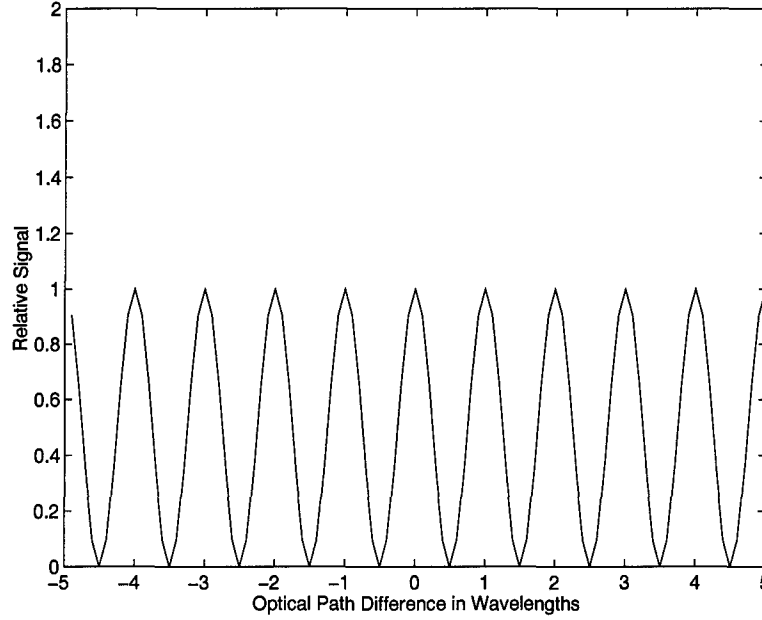


Figure 2. Interferogram obtained from a monochromatic source. Intensity peaks equally spaced one wavelength in optical path difference (OPD). Zero path point refers to condition when OPD equals zero.

can occur, causing the light intensity on the detector to decrease. If one of the mirrors is moved at a constant velocity, the interference pattern alternates between constructive and destructive interference giving rise to an interferogram, Fig. 2, which for monochromatic light has the form of a cosine fluctuation(9:1024A). When the path lengths in the two arms of the interferometer are equal, we refer to this condition as the zero path point. For a monochromatic source, the interferogram can be written as

$$I(x) = B(\sigma) [1 + \cos(2\pi\sigma x)] \quad (1)$$

where I is the intensity on the detector, x is the OPD in cm, $2B$ is the intensity of the source, and σ is the wavenumber of the radiation in cm^{-1} (9:1024A).

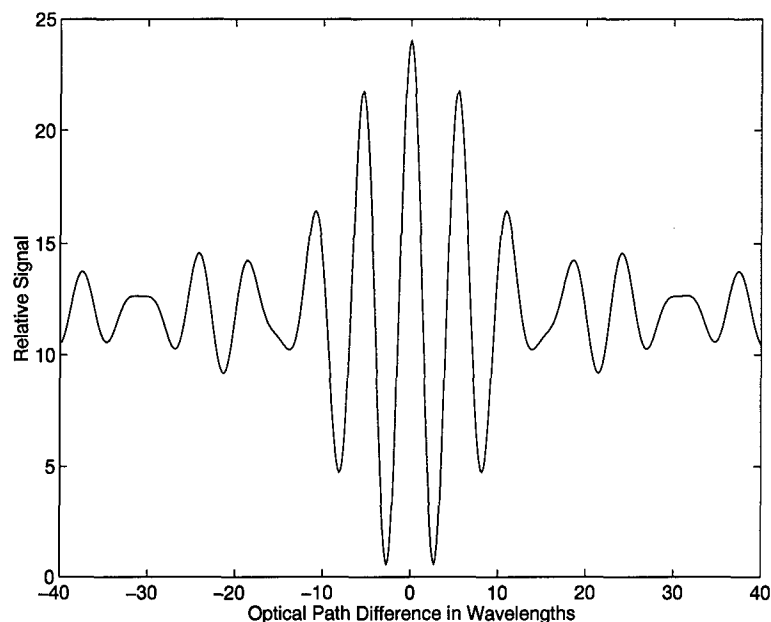


Figure 3. Interferogram obtained from a polychromatic source. Central intensity peak result of all light wavelengths being in phase at zero path point. Extension of waveform in both directions eventually yields a constant level.

For a polychromatic source, the detector sees the sum of the interference for all wavelengths (9:1024A). The interferogram takes the form as shown in Fig. 3. At the zeroth path point, a central fringe, or intensity peak, is observed due to the constructive interference of all wavelengths of the source. As the OPD increases, the envelope of the interferogram decreases until a steady-state *dc* level is reached (9:1024A). The decrease in the envelope of the interferogram can be viewed in two equivalent ways, as either a “dephasing of elementary fringes, or in terms of a loss of correlation due to the finite pathlength delay” (13:160). The modulation of the fringe drops to approximately zero when the OPD increases beyond the coherence length of the source (13:163). The intensity distribution near the central portion of the fringe contains low resolution information about the source, while the intensity at

higher OPDs contain high resolution information. The detector used in FTS must have adequate dynamic range to ensure that both the high and low resolution information is measured, and the source must remain stable during the observation time, or the time that the interferogram is being sampled (9:1024A).

Bell showed that the interferogram is the autocorrelation of the incident wave amplitude (2:43) and the Wiener-Khinchine theorem states that the Fourier transform of the autocorrelation of a function is the power spectrum for that function (19:8). Hence, the interferogram and the power spectrum are Fourier transform pairs (21:59) (33) (27:19) (32:250).

To record the entire interferogram produced by the source, the mirrors in an ideal Michelson interferometer would traverse from an OPD of zero to an OPD of infinity. Since this is not realizable in practical applications, the sampled interferogram is limited to OPDs from zero to some maximum value. This is equivalent to multiplying the interferogram by a rectangular "window" function. The window function being zero outside the measured region (24:167).

$$I(x) = B(\sigma) [1 + \cos(2\pi\sigma x)] \text{rect}\left(\frac{x}{d}\right) \quad (2)$$

where the "window" function is given as

$$\text{rect}\left(\frac{x}{d}\right) = \begin{cases} 1, & -d/2 < x < d/2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Taking the Fourier transform of the measured interferogram to yield the desired spectra will result in sidelobes due to the window function (8:68) (24:167). The Fourier transform of

Eqn. (3) is a function commonly referred to as the *sinc* function. It consists of a central lobe with decreasing sidelobes. A common method to reduce these sidelobes is to multiply the interferogram by an apodizing function (8:68) (32:252) (2:51).

Errors in FTS spectral recovery are due to the aperture effect, tilt and aberrations, truncation, phase and compensation errors, and noise (19:14). These errors can be eliminated or reduced to a point that the computed spectrum closely approximates that of the source, especially when high signal-to-noise ratios are present. The primary sources of noise are detector noise, photon noise, scintillation noise, and digitization noise. Detector noise is inherent in the detection mechanism itself and is independent of the incident signal level. Photon noise occurs because of the random arrival times of photo-events in detectors, usually modelled mathematically as a Poisson process. Scintillation noise refers to a slow drift in the intensity of the light incident on the interferometer input. Several techniques have been used to minimize effects of scintillation. Digitization noise refers to random fluctuations caused by the analog-to-digital conversion process and can be reduced with sophisticated electronics (27:28-29).

Sakai stated that the signal-to-noise ratio in the recovered spectrum, not that in the interferogram, is the prime factor that determines the quality of a measurement in FTS (27:19). Goodman developed an expression for the SNR of a Michelson interferometer (13). He used the Discrete Fourier Transform (DFT) of the count vector as an estimation tool to estimate the fringe parameters. From his analysis, he expressed the SNR in fringe amplitude

by

$$\left(\frac{S}{N}\right)_{rms} = \sqrt{\frac{K_1 + K_2}{2}} \vartheta \quad (4)$$

where ϑ is the visibility in fringe amplitude and $K_1 = \alpha A \tau N I_1$ and $K_2 = \alpha A \tau N I_2$ are the average number of photo-events generated by the interfering beams across the entire array (13:490-500).

2.2 Introduction to the Sagnac Interferometer

The Sagnac interferometer was first introduced by Sagnac to measure the rotation of the earth (32:183) and is commonly used today in navigation gyros (6). The Sagnac interferometer, Fig. 4, has advantages that are desirable for applications in Space Object Identification (SOI). For instance, the Sagnac interferometer has no mechanical moving parts, is easy to align and is very insensitive to equipment vibrations or mirror displacements (30:269). Also, the Sagnac interferometer has an advantage that allows it to measure the interferogram of rotating or unstable spacecraft.

Unlike the Michelson interferometer which uses a single detector and requires the interferogram to be scanned during the observation interval, the Sagnac interferometer spatially distributes the interferogram across a linear array of detector elements, thereby allowing simultaneous sampling of the entire interferogram (27:4222). In the Sagnac interferometer, the incident collimated wavefront is split into two beams by a beam-splitter. The beams follow the same paths in opposing directions, are reflected by the mirrors, and are recombined by the beam-splitter. One of the mirrors is translated such that the recombined beams are lat-

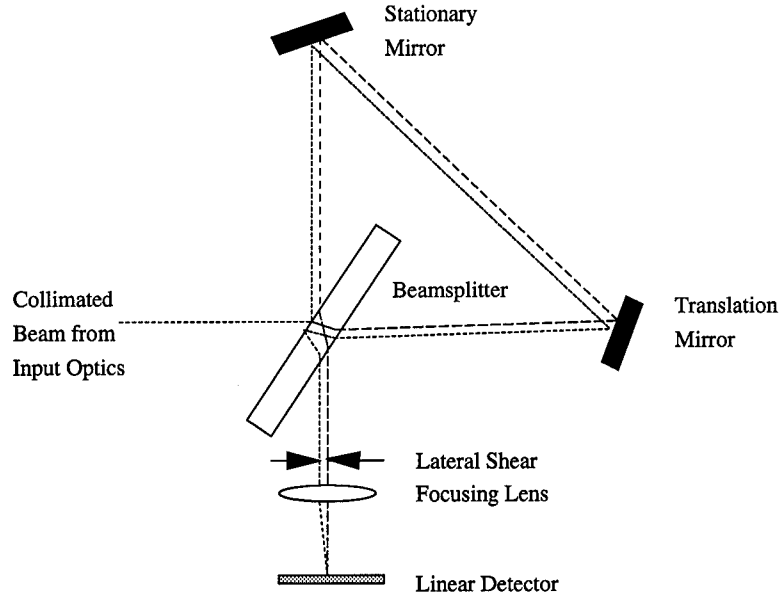


Figure 4. Schematic Diagram of the Sagnac Interferometer

erally displaced, or sheared. The magnitude of the lateral shear is controlled by the position of the mirror and the wavelength of the source. The mirror position is set during alignment and is stationary during the observation interval. Therefore, during the observation, the lateral shear is controlled only by the wavelength of the source. Each source location, or emitter, provides an irradiance that, when combined with the irradiance of the other source emitters, yield a spatially coherent interferogram on the detector plane. The spherical lens in Fig. 4 focuses the interferogram onto the detector.

Okamoto *et al.* modeled the recorded interferogram by the expression

$$\kappa(m) = \int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) b(\sigma) \left(1 + \cos \left[\frac{2\pi\sigma l(\sigma)(dm - \phi)}{f} \right] \right) d\sigma \quad (5)$$

where $s(\sigma)$ is the spectral distribution of the source, $b(\sigma)$ is the system transfer function, d is the separation between detector elements in the linear array, ϕ is the distance between the center of the optical axis and the cell for zero OPD, f is the focal length of the focusing lens, σ_{min} and σ_{max} are the minimum and the maximum wavenumbers within the bandpass of the detector or optical filter, and l is the lateral shear (or distance between virtual sources) (31:4222).

2.3 SNR of the Sagnac Interferometer

Assume that the interferometer is well designed and aligned so that $\phi = 0$ for all $\sigma_{min} < \sigma < \sigma_{max}$. Also, assume l constant and $b(\sigma) = 1$ throughout bandpass. Now, Eqn. (5) becomes

$$\kappa(m) = \int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) \left[1 + \cos\left(\frac{2\pi\sigma l m d}{f}\right) \right] d\sigma \quad (6)$$

Or, rewritten as

$$\kappa(m) = \int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) d\sigma + \int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) \cos\left(\frac{2\pi\sigma l m d}{f}\right) d\sigma \quad (7)$$

Following Goodman's approach (13:494), the Discrete Fourier Transform (DFT) of the count vector Eqn. (7) must be obtained in order to estimate the spectral SNR of the measurement. The DFT of Eqn. (7) is

$$\begin{aligned} K(p) &= \frac{1}{N} \sum_{n=0}^{N-1} \kappa(m) e^{-j\frac{2\pi m p}{N}} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \left[\int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) d\sigma + \int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) \cos\left(\frac{2\pi\sigma l m d}{f}\right) d\sigma \right] e^{-j\frac{2\pi m p}{N}} \end{aligned} \quad (8)$$

where $K(p)$ represents the DFT of $\kappa(m)$.

The first term in the right hand side of Eqn. (7) is the total light within the bandpass of the system and is determined by radiometric calculations. Here, this light is assumed to be a constant value given by

$$K(0) = \int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) d\sigma \quad (9)$$

Given the definition of the normalized Power Spectral Density (PSD) (13:164), and assuming that $s(\sigma)$ is zero outside the bandpass measured by the detector, the normalized PSD can be written as

$$\hat{G}(\sigma) = \frac{s(\sigma)}{\int_{\sigma_{min}}^{\sigma_{max}} s(\sigma) d\sigma} \quad (10)$$

and Eqn. (8) becomes

$$K(p) = \frac{1}{N} \sum_{n=0}^{N-1} \left[K(0) + K(0) \int_{\sigma_{min}}^{\sigma_{max}} \hat{G}(\sigma) \cos\left(\frac{2\pi\sigma lmd}{f}\right) d\sigma \right] e^{\frac{-j2\pi np}{N}} \quad (11)$$

Goodman states that when the two interfering beams are equal in intensity, the visibility is the Fourier transform of the normalized PSD (13:162-164)

$$\vartheta(\tau) = \int_{\sigma_{min}}^{\sigma_{max}} \hat{G}(\sigma) \cos(2\pi\sigma\tau) d\sigma \quad (12)$$

where $\tau = lmd/f$. Therefore, $\vartheta(\tau)$ can be substituted into Eqn. (11) to give

$$\begin{aligned} K(p) &= \frac{1}{N} \sum_{n=0}^{N-1} [K(0) + K(0) \vartheta(\tau)] e^{\frac{-j2\pi np}{N}} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} K(0) e^{\frac{-j2\pi np}{N}} + \frac{1}{N} \sum_{n=0}^{N-1} K(0) \vartheta(\tau) e^{\frac{-j2\pi np}{N}} \end{aligned}$$

$$= K(0) \frac{1}{N} \sum_{n=0}^{N-1} e^{-\frac{j2\pi np}{N}} + K(0) \frac{1}{N} \sum_{n=0}^{N-1} \vartheta(\tau) e^{-\frac{j2\pi np}{N}} \quad (13)$$

Also note that

$$\frac{1}{N} \sum_{n=0}^{N-1} \vartheta(\tau) e^{-\frac{j2\pi np}{N}} = \hat{G}(p) \quad (14)$$

Substituting Eqn. (14) into Eqn. (13) gives

$$K(p) = K(0) \delta(p) + K(0) \hat{G}(p) \quad (15)$$

where $\delta(p)$ is the Dirac delta function, with magnitude of one when $p = 0$ and magnitude of zero elsewhere. Okamoto *et al* give the bin p to wavelength conversion as $\lambda = Nld/fp$ (30:271).

Eqn. (15) shows that the DFT of the interferogram detected by the Sagnac interferometer yields a *dc* component, and an *ac* component determined by the normalized power spectral density (PSD) of the source, $\hat{G}(p)$. For spectral SNR analysis, we are only concerned with the SNR at a particular wavelength. Therefore, the *dc* term is ignored and Eqn. (15) becomes

$$K(p) = K(0) \hat{G}(p) \quad (16)$$

The general SNR expression, for the case where the signal obeys Poisson statistics and the noise is additive zero-mean Gaussian, is given as (14:2-10)

$$SNR = \frac{\kappa_s}{\sqrt{\kappa_s + \kappa_b + N\sigma_{rn}^2}} \quad (17)$$

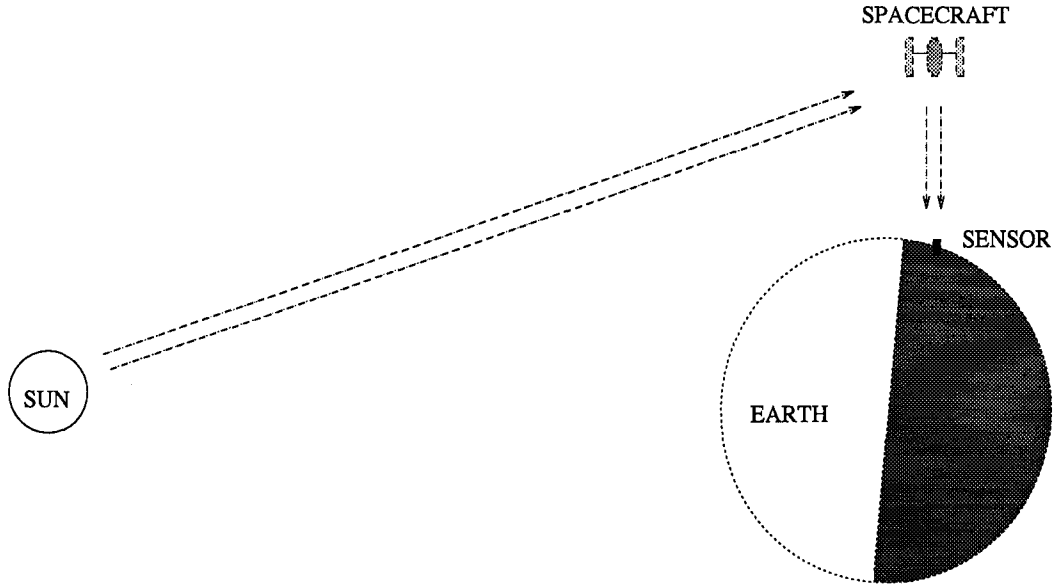


Figure 5. Satellite Observation Scenario. The telescope observes the light reflected from the satellite surface materials.

where κ_s is the detected signal count, κ_b is the detected background count, N is the number of detector elements in the CCD array, and σ_{rn} is the read noise of a single CCD detector element.

To further develop the Sagnac SNR expression, a relationship between κ_s and $K(p)$ must be found. Hrovat (14) studied the radiometry for the satellite scenario under investigation, see Fig. 5. The satellite is illuminated by the sun, and the telescope collects the reflected light. In Hrovat's analysis, he assumed that the light passed through a spectral filter prior to being sensed by the detector. He allowed the bandpass of the spectral filter to be 5, 10, or 20 nm. The results of the Sagnac SNR derivation, Eqn. (16), show that Hrovat's approach can be directly applied to this problem. The following is a brief summary of Hrovat's derivation.

Table 1. Values for well-known constants rounded to the value generally used in practice.

Constant	Description	Value	Units
h	Planck's constant	6.625×10^{-34}	Joules-Second
c	Speed of light in vacuum	3×10^8	meters/second
k	Boltzmann's constant	1.38×10^{-23}	Joules/Kelvin

First, he showed that the detected count from the background κ_b is negligible when compared to the detected count from the satellite κ_s , reducing Eqn. (17) to

$$SNR = \frac{\kappa_s}{\sqrt{\kappa_s + N\sigma_{rn}^2}} \quad (18)$$

The spectral irradiance on the satellite, assuming the sun is the sole source of illumination, is given by

$$E_{source}(\lambda) = M_{bb}(\lambda) \left(\frac{R_{sun}}{r_{sun}} \right)^2 \cos(\theta_2) \quad (19)$$

where R_{sun} is the radius of the sun, r_{sun} is the range, or distance, from the sun to the satellite, θ_2 is the angle, as seen by the spacecraft, between the direction of the sun and the direction of the satellite surface normal, and $M_{bb}(\lambda)$ is the blackbody exitance of the sun and is given by Planck's law (4:54)

$$M_{bb}(\lambda) = \frac{2\pi hc^2}{\lambda^5 \left(e^{\frac{hc}{\lambda k_B T}} - 1 \right)} \quad (20)$$

where h , c , and k_B are familiar constants shown in Table 1. T is the temperature of the sun, here assumed to be 5770 K.

Assuming the satellite to be a Lambertian scatterer (4:20), the spectral irradiance at the sensor aperture is given by

$$E_T(\lambda) = \frac{A_T \cos(\theta_1)}{\pi R^2} \rho(\lambda) \tau_a(\lambda) E_{source}(\lambda) \quad (21)$$

where A_T is the surface area of the satellite, θ_1 is the angle between the direction of the telescope and the direction of the satellite surface normal, τ_a is the transmission of the atmosphere, $\rho(\lambda)$ is the spectral reflectivity of the surface material on the satellite, and R is the distance, or range, from the telescope to the satellite.

The total power on the detector is the product

$$\Phi_e(\lambda) = A_o \tau_o(\lambda) E_T(\lambda) \quad (22)$$

where A_o is the area of the telescope, and τ_o is the optical transmission. The irradiance on the detector is the ratio of the total power on the detector to the total area of the detector

$$E(\lambda) = \frac{\Phi_e(\lambda)}{A_d} \quad (23)$$

which is physically the same as the quantity expressed in Eqn. (16)

$$E(\lambda) = K(0) \hat{G}(\lambda) \quad (24)$$

where the functional dependence on λ was included by making the substitution $p = Nld/f\lambda$.

Goodman converts the irradiance into an electron count with the expression

$$\kappa_s(\lambda) = \alpha A_d t_d E(\lambda) \quad (25)$$

where $\alpha = \eta(\lambda) \lambda / hc$, $\eta(\lambda)$ is the detector quantum efficiency, t_d is the detector integration time, and $E(\lambda)$ is given either by Eqn. (23) or Eqn. (24) (13:493). Combining the above equations gives the following expression for the signal count on the detector at wavelength λ

$$\kappa_s(\lambda) = \frac{\lambda \eta(\lambda)}{hc} A_d t_d \frac{A_o \tau_o(\lambda)}{A_d} \frac{A_T \cos \theta_1}{\pi R^2} \rho(\lambda) \tau_a(\lambda) M_{bb}(\lambda) \left(\frac{R_{sun}}{r_{sun}} \right)^2 \cos \theta_2 \quad (26)$$

which reduces to

$$\kappa_s(\lambda) = \frac{2 t_d A_o A_T \cos \theta_1 \cos \theta_2 R_{sun}^2}{R^2 r_{sun}^2} \frac{\tau_o(\lambda) \tau_a(\lambda) \rho(\lambda) \eta(\lambda)}{\lambda^4 \left(e^{\frac{hc}{B^T}} - 1 \right)} \quad (27)$$

with the substitution of Eqn. (20) for $M_{bb}(\lambda)$. Therefore, for the satellite observation scenario in Fig. 5, the spectral SNR of the Sagnac interferometer is given by

$$SNR(\lambda) = \frac{\kappa_s(\lambda)}{\sqrt{\kappa_s(\lambda) + N \sigma_{rn}^2}} \quad (28)$$

where $\kappa_s(\lambda)$ is as given in Eqn. (27).

Table 2. Values chosen for radiometric models. These values are typical of those seen in practice.

Variable	Description	Designated Value	Units
A_o	Area of optic; 1.6 m dia.	2.0106	m^2
θ_1	Satellite-sensor angle	80	deg
θ_2	Satellite-sun angle	80	deg
R_{sun}	Radius of the sun	6.96×10^8	m
r_{sun}	Range from sun to satellite	1.5×10^{11}	m
τ_o	Optical transmission	0.4	N/A
τ_a	Atmospheric transmission	0.8	N/A
ρ	Spectral reflectivity	0.5	N/A
η	Detector QE	0.4	N/A
T	Sun blackbody temperature	5770	K
N	Number of detectors in Array	1024	N/A
σ_{rn}	Detector read-noise	12	e^{-1}

2.4 Expected SNRs using Radiometric Models

This section shows the results of applying Eqn. (27) and Eqn. (28) to possible satellite observation scenarios. The results illustrate the range of possible signal-to-noise ratios for observation of satellites using the Sagnac interferometer.

To simplify the analysis, but yet maintain realistic results, many variables in Eqn. (27) were set to specified values as shown in Table 2. Note that the functional dependencies on λ for τ_o , τ_a , ρ , and η were removed and $1/\lambda^4 (e^{hc/\lambda k_B T} - 1)$ was evaluated over the wavelength range under consideration, 300 nm to 1 μm . The constant values shown in Table 2 were chosen under the conditions that they were representative of the mean value expected in practice, and that they provided conservative estimates of the SNR when inserted into Eqn. (27) and Eqn. (28). The remaining variables, t_d , A_T , and R , were chosen in order to simulate realistic scenarios and to illustrate their influence on the SNR.

Table 3. Radiometric scenarios and results. The profile refers to the size of the satellite, small or large, and the range to the satellite, LEO, U-LEO, MEO, and GEO. A_T is the total satellite surface area, and t_d is the detector integration time. The SNR refers to the average SNR across the 300-1000 nm band.

Profile	Range (km)	$A_T (m^2)$	$t_d (s)$	SNR
Small LEO	1000	1	.01	28.13
Large LEO	1000	50	.01	274.67
Small U-LEO	5000	1	.10	5.06
Large U-LEO	5000	50	.10	144.31
Small MEO	20000	1	10	3.23
Large MEO	20000	50	10	151.35
Small GEO	40000	1	300	4.43
Large GEO	40000	50	300	217.56

The scenarios that were modeled are shown in Table 3. The values for the variables in Table 3 were chosen with the following considerations in mind. The ranges R are typical of distances between satellite and ground-based telescope for low-earth orbiting (LEO), upper low-earth orbiting (U-LEO), mid earth orbiting (MEO), and geosynchronous (GEO) satellites. The satellite surface areas A_T are realistic for both small and large satellites. The detector integration times t_d were chosen with two considerations in mind: the result of the observation, image or spectrum, and the resulting SNR. For observations at short ranges, the performance of the Sagnac interferometer is suitable for imaging purposes, and t_d is short such that atmospheric scintillation effects are reduced. At longer ranges, the imaging capabilities of the Sagnac interferometer degrades and t_d is increased to provide sufficient SNRs for spectrum measurements. The resulting average SNRs, shown in Table 3, are in an approximate interval of 1 to 100. These SNRs are representative of those obtained by an ideal Sagnac interferometer that measures the spectrum of an orbiting spacecraft.

2.5 *Summary*

This chapter included an introduction to Fourier Transform Spectroscopy and the Sagnac interferometer. The derivation of the SNR expression for the Sagnac interferometer led to an expression, Eqn. (16), previously modelled by Hrovat (14) for ground-based observations of orbiting spacecraft. Therefore, Hrovat's radiometric analysis is applicable to this research and is used to estimate the expected SNRs for data obtained from the Sagnac interferometer, Table 3. The resulting SNRs are in the range of 1 to 300. The following chapters provide percent composition estimates on data having average SNRs of 1, 10, and 100.

III. Theory for Estimation of Percent Composition

Satellite surfaces are generally composed of many materials, with each material having its own characteristic spectrum. In the scenario under investigation, Fig. 5, the Sagnac interferometer measures the spectrum of the satellite. It is assumed throughout this effort that the sensor in the Sagnac interferometer measures the solar illumination reflected by the satellite in the wavelength range, 300 nm to 1 μm . Each detector in the sensor's linear array records the irradiance of the light in its associated spectral band.

The percent composition refers to the percentage of the total satellite surface area that can be attributed to a single material, where as stated in Section 1.3 spectral mixing is assumed to behave linearly. Eqn. (27) showed that the electron count $\kappa_s(\lambda)$ is directly related to the satellite surface area A_T and the spectral reflectivity $\rho(\lambda)$ of the material. Defining a new variable $\gamma(\lambda)$ as the product of all variables in Eqn. (27) other than A_T and $\rho(\lambda)$ gives

$$\kappa_s(\lambda) = \gamma(\lambda) A_T \rho(\lambda) \quad (29)$$

for the electron count (13). If the satellite's surface area is composed of two materials, the count may be determined from

$$\kappa_s(\lambda) = \gamma(\lambda) (A_{T1}\rho_1(\lambda) + A_{T2}\rho_2(\lambda)) \quad (30)$$

where A_{T1} and $\rho_1(\lambda)$ are the surface area and spectral reflectivity for the first material, and A_{T2} and $\rho_2(\lambda)$ are the surface area and spectral reflectivity for the second material. By

estimating $\gamma(\lambda)$ from atmospheric and radiometric models and then normalizing Eqn. (30) by $\gamma(\lambda)$ and the total surface area, the normalized spectrum can be given by the expression

$$b(\lambda) = x_1 \rho_1(\lambda) + x_2 \rho_2(\lambda) + n(\lambda) \quad (31)$$

where x_1 and x_2 are the percent compositions for material 1 and material 2, respectively, and $n(\lambda)$ represents the noise.

Eqn. (31) can be generalized to more than two materials and rewritten in matrix form as

$$\vec{b} = A\vec{x} + \vec{n} \quad (32)$$

where A now represents an L -by- M matrix whose columns contain known spectral reflectivities of satellite materials, hereafter referred to as the material database. L is the number of spectral bins and M is the number of satellite materials whose spectra are present in the material database. \vec{x} is a M -by-1 vector representing the percent compositions for the $r \leq M$ materials that contributed to the observation \vec{b} . This is now cast as an inverse problem—given \vec{b} , find \vec{x} and evaluate the errors.

As mentioned previously, the objective of this research is to determine if the observation resulting from the interferogram recorded by a Sagnac interferometer can be used in a system that provides reliable estimates of the spacecraft's material composition. Three systems for estimation of percent composition, or spectral unmixing, are proposed in this chapter along with the theory for their implementation. The sections in this chapter include:

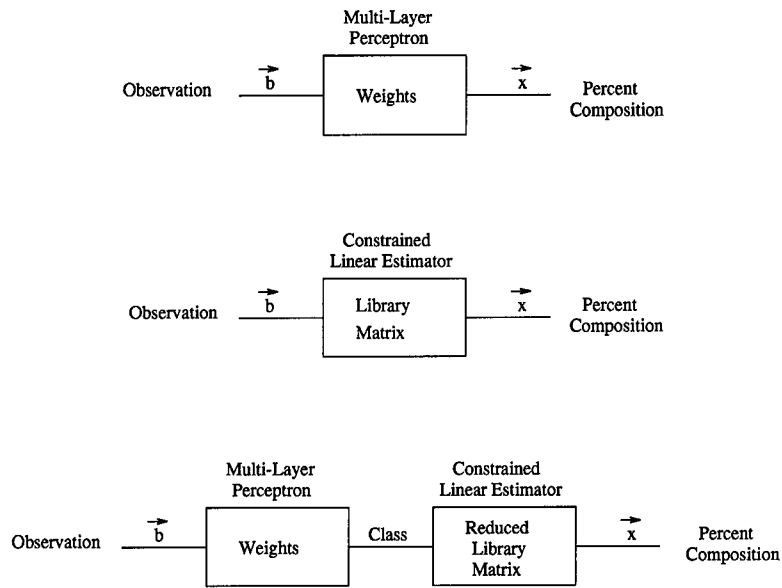


Figure 6. Block Diagram of Three Proposed Systems

- Introduction to the System Architectures
- Concepts for Material Classification
- Concepts for Constrained Linear Estimation

3.1 Introduction to the System Architectures

The block diagrams of the three systems for estimation of material compositions are shown in Fig. 6. The first system uses a MLP to directly estimate the percent composition, the second system uses a constrained least squares estimator, and the third system uses a hybrid of both the MLP and the constrained least-square linear estimator in an attempt to reduce the root-mean-square (RMS) error in the estimate. The remainder of this chapter provides an introduction to the theory necessary for implementation of these systems.

3.2 Concepts for Material Classification

Schalkoff defines pattern recognition as “the science that concerns the description or classification (recognition) of measurements (28:2).” A common problem encountered in pattern recognition is classifier design. A classifier labels a measurement into one or more classes based upon some statistical, structural, or neural decision rule (28:4).

As an example, a problem encountered in Automatic Target Recognition (ATR) is the design of a system that automatically distinguishes objects, such as tanks and jeeps. To distinguish tanks from jeeps with high reliability may require that the length, height, and weight be measured for each object and stored in a vector. This vector is referred to as a feature vector, where each measurement is one of its elements. To train the classifier, the feature vectors for the tank are compared with the feature vectors for the jeep. The data used in this training stage is referred to as “training data”, and requires that each feature vector be labeled. After the classifier has been designed, it is tested with “test data” using either a statistical, structural, or neural decision rule. The test data is independent of, but similar to, the training data. The classifier assigns each test vector to the class, tank or jeep, upon which it is most similar (28:14), without referring to the test vector’s actual class. Following classification of all test samples, the class assigned by the classifier to each sample is compared to its actual class label to determine the test error rate.

The ability of a classifier to discriminate between classes is often apparent by visualizing the feature vectors in feature space (28:13). In feature space, the coordinate for each axis is given by an element of the feature vector. Fig. 7 shows a two-dimensional feature space

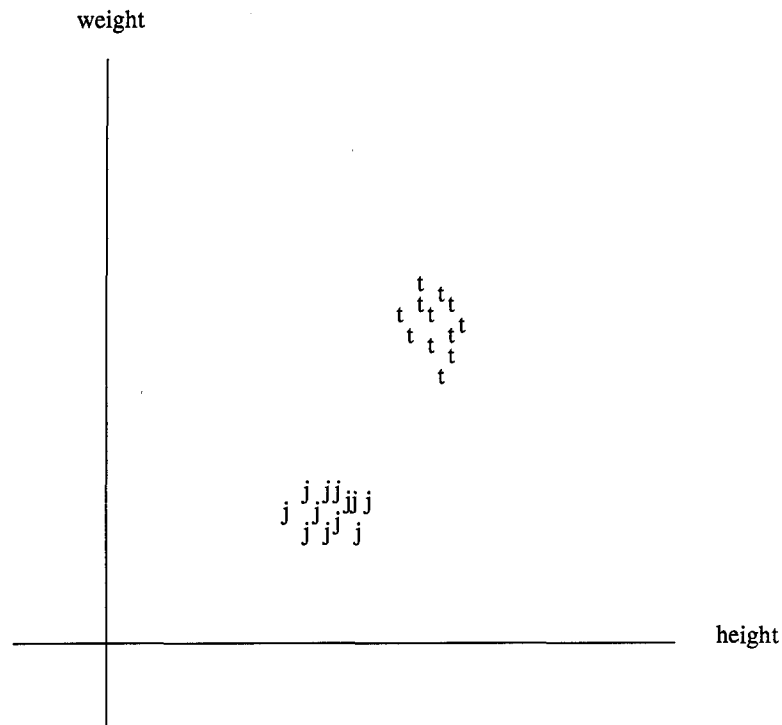


Figure 7. Feature Space for Tank and Jeep Example

for the tank and jeep example, where the values for two of the features, weight and height, are plotted along the axes. Based upon these features, the data points for the jeep (j) are clearly separated from the data points for the tank (t), and a classifier should have little difficulty separating the two classes. The features for most problems in pattern recognition are more difficult to separate. In classification problems of greater difficulty, a classifier distinguishes classes by partitioning the feature space into decision regions (28:15), where each of the c possible classes is concentrated in its own specific region (unimodal distribution) or regions (multimodal distribution). To aid the classifier in its performance, the designed system chooses features such that features from data within the same class have minimum covariance, while features from data in different classes have maximum covariance (28:92).

3.2.1 *Feature Extraction and Data Analysis.* The features are chosen, or extracted, using methods commonly presented in pattern recognition literature. The term “feature extraction” refers to the process of extracting a lower dimensional feature vector from the raw data that continues to represent the essence of the data (17:198). As stated in the previous section, the extracted features should have minimum variation with data of the same class and maximum variation with data of different classes. In this thesis, two approaches were initially used to extract the features: Fisher discriminants, and principal component analysis (PCA). The intermediate results illustrated that the Fisher discriminant method for feature extraction yielded lower overall classification errors, and was therefore the chosen method for feature extraction. The PCA method was thereafter used solely for data analysis.

3.2.1.1 *Fisher Discriminants.* The Fisher discriminants use the concept that the degree of overlap in feature space between different classes is proportional to the distance between their distributions and inversely proportional to their scatter,

$$f = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (33)$$

where m_1 and m_2 are the mean value of the feature for the classes, and σ_1^2 and σ_2^2 are the variance of the feature for each class (22:177). In order to reduce the number of features presented to the classifier, the Fisher discriminants are calculated for each feature in the feature vector and then sorted from high to low value, the features associated with the highest discriminants are retained and presented to the classifier (28:90). Although this procedure for feature extraction is straightforward, it does not guarantee that the retained

set of features are best for classification purposes. Fisher discriminants are discussed further in Section 4.2.2.

3.2.1.2 Principal Component Analysis. The principal components of a data set represent the selection of a coordinate system whose axes are along the directions of maximum variance in the data (18:362). These directions are obtained by computing the eigenvectors of the data's covariance matrix. The magnitude of the eigenvalue corresponding to an eigenvector represents the variance in the data along the direction of that eigenvector. For purposes of visual analysis, the data is projected into the plane of eigenvectors corresponding to the two largest eigenvalues. This plane is called the principal component plane, or simply the principal plane. Principal components are discussed further in Section 4.2.1.

3.2.2 The Parzen Classifier. A necessary assumption for the Parzen classifier is that the decision regions be defined in probabilistic terms, and that all relevant probabilities be known. The implementation of this classifier in the literature generally refers to the work of Duda and Hart (7). They showed that if $P(\omega_j)$ is the *a priori* probability that nature is in state ω_j , then the *a posteriori* probability $P(\omega_j|x)$, interpreted as the conditional probability of being in class ω_j given the observation x , can be computed from the conditional probability density function (PDF) $p(x|\omega_j)$ by Bayes rule:

$$P(\omega_j|x) = \frac{p(x|\omega_j) P(\omega_j)}{p(x)} \quad (34)$$

where

$$p(x) = \sum_{j=1}^s p(x|\omega_j) P(\omega_j) \quad (35)$$

is the total probability of observation x , and s is the total number of states.

The Bayes approach to classification is one of finding the $P(\omega_j)$ that minimizes the overall error given by the expression (7:13)

$$p(error) = \int_{-\infty}^{+\infty} P(error|x) p(x) dx \quad (36)$$

where $P(error|x)$ is the probability of error given the observation x .

For a c class problem, the classifier minimizes the error by assigning an unknown test vector x to the class ω_i having the largest discriminant function (7:17). The discriminant function is calculated for each of c classes by using the expression (7:18)

$$g_i(x) = \frac{p(x|\omega_i) P(\omega_i)}{\sum_{j=1}^c p(x|\omega_j) P(\omega_j)} \quad (37)$$

If $P(\omega_j)$ is assumed equal for all classes, then Eqn. (37) reduces to

$$g_i(x) = \frac{p(x|\omega_i)}{\sum_{j=1}^c p(x|\omega_j)} \quad (38)$$

or $g_i(x)$ equals the normalized conditional PDF for its associated class. In determining class assignment for the test vector, if

$$g_i(x) > g_j(x) \quad (39)$$

for all $j \neq i$ then assign x to class ω_i .

In practice, $P(\omega_i)$ and $p(x|\omega_i)$ are not perfectly known, and must be obtained from data. The literature (28:58) refers to two approaches for estimating $p(x|\omega_i)$, based on whether a functional form for the conditional PDF is assumed. Here, a nonparametric Parzen Window approach (28:70) is taken where the functional form for the conditional PDF is not assumed. Parzen introduced the Parzen window approach in 1961 (23). The Parzen estimate of the conditional PDF for class ω_i given the observation x in some n -dimensional space (R^n) is given by the expression (12:634)

$$\hat{p}_i(x) = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{1}{h^n} k_i \left(\frac{x - x_j^{(i)}}{h} \right) \quad (40)$$

where $x_j^{(i)}$ represents each of the N_i samples in class ω_i , $k_i(\cdot)$ is a window function with a volume of one in the n -dimensional set of real numbers R^n , and h controls the spread of k_i . The Parzen approach can be thought of as centering a unit volume window around the observation vector x and using the percentage of the N_i samples that fall within the window as an estimate of the conditional PDF. A commonly used window function is the Gaussian function (7:23) written as

$$f(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{\mu})^t \Sigma^{-1} (\vec{x} - \vec{\mu}) \right] \quad (41)$$

where $\vec{\mu}$ is a vector of the means for all \vec{x} and Σ is the n -by- n covariance matrix. When used as a Parzen window, the Gaussian function takes the form (20:32)

$$k_i \left(\frac{x - x_j^{(i)}}{h} \right) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_i|^{\frac{1}{2}} h^n} \exp \left[-\frac{1}{2h^2} (x - x_j^{(i)})^t \Sigma_i^{-1} (x - x_j^{(i)}) \right] \quad (42)$$

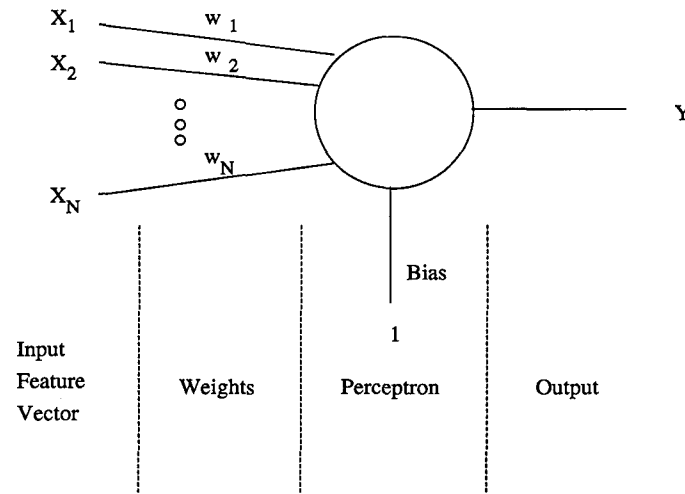


Figure 8. Input/Output Diagram of a Single Perceptron (25:55)

and can be directly inserted into Eqn. (40) to estimate $p_i(x)$.

3.2.3 The Multi-Layer Perceptron. Multi-Layer Perceptrons (MLP) are commonly used as pattern classifiers in pattern recognition problems. A complete introductory treatment on MLPs is beyond the scope of this thesis, and can be found in the literature (25) (28).

A MLP is a network of single perceptrons, Fig. 8, interconnected in such a way to efficiently classify input patterns. Each perceptron, introduced by Rosenblatt (26), separates the feature space into two regions in a nonlinear manner. It does this in two stages. First, it multiplies each of its inputs by a weight on the interconnection associated with that input and computes a weighted sum over all inputs (a bias term is here assumed to be an input

with weight of one and is included in the weighted sum, as shown in Fig. 8)

$$x_j = \sum_i y_i w_{ij} \quad (43)$$

where x_j is the total weighted input or actuation, y_i is the output of the i th perceptron in the previous layer, and w_{ij} is the weight of the connection between the i th and j th perceptrons. Then, the perceptron uses a nonlinear function to transform the input into a single output activity (25:55). Two nonlinear functions were used in this research, the sigmoid function

$$y_j = \frac{1}{1 + \exp^{-x_j}}$$

and the hyperbolic tangent function

$$y_j = \frac{e^{x_j} - e^{-x_j}}{e^{x_j} + e^{-x_j}} \quad (44)$$

shown in Fig. 9.

An MLP solves pattern recognition problems that require more than a linear decision boundary in order to separate the data. The manner in which the perceptrons are interconnected are a topic of continual study. Cybenko showed that a three layer MLP provides the needed complexity to solve any classification problem, given enough perceptrons in the hidden layer (5). A three layer MLP consists of an input layer, a "hidden" layer, and an output layer, Fig. 10. The input layer is simply a location for the data to be temporarily

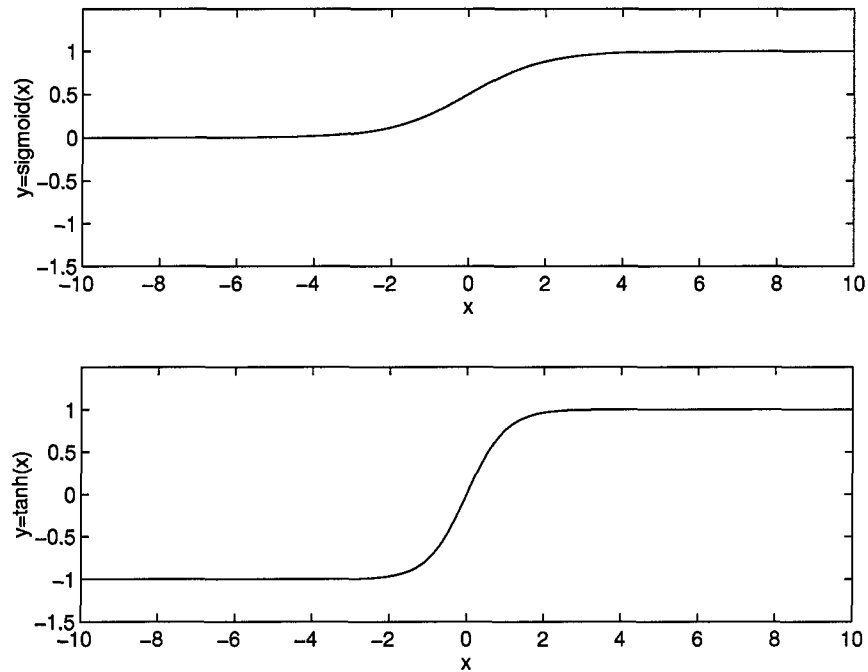


Figure 9. Sigmoid and Hyperbolic Tangent Functions

stored prior to being entered into the “hidden” layer. The “hidden” and output layers have perceptrons that perform the transformations according to Eqn. (43) and Eqn. (44).

There are three primary concerns in the design of a MLP: the size of the network, the time required to learn the decision boundaries, and the ability to generalize on data outside the training set (15:16). Brief comments on network size and generalization are presented in the following.

In a three-layer MLP, the size of the network refers to the total number of nodes in the three layers. Choosing the network size is important. If the network is too small, the MLP will not solve the problem, and if the network is too large, the MLP will oversolve the problem (i.e. the weights will learn the training data), and provide poor generalization.

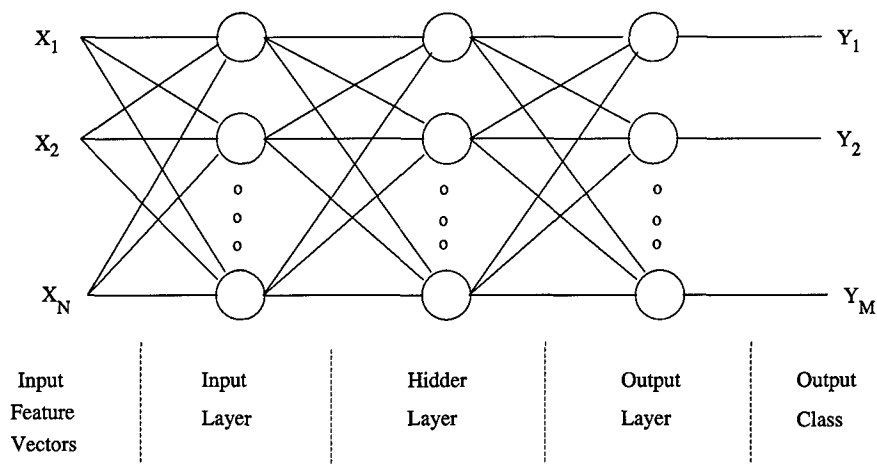


Figure 10. Interconnection Diagram for a Multi-Layer Perceptron (25:62)

Poor generalization refers to the situation where the classifier performs well on the training data, but does not perform well on the test data. For a three-layer MLP, the number of input nodes and the number of output nodes are defined by the pattern recognition problem itself. Each input node is associated with a single element in the feature vector, and each output node is associated with an output class assignment. No approach has been found that chooses the optimal number of hidden nodes for a given problem (15:16), but useful guidelines have been found that put upper limits on this number (29) (1). Huang gives an upper bound on the number of hidden nodes as the number of training samples entered into the network (29), and Baum confines this further by stating that, for greater than 90% classification accuracy, the number of weights in the network should be less than one-tenth the number of training samples (1:153).

The ability of a classifier to generalize is of concern because the ultimate goal of the classifier is to properly classify measurements that have not been previously tagged. Foley

demonstrated that for valid generalization there should be no more than three times as many training samples per class as there are features within the feature vector (11). Baum and Haussler (1) claim that the number of training samples for valid generalization is given by W/ϵ , where W is the size of the network and ϵ is the desired error rate on the test data.

After the neural network architecture has been chosen, the weights in the network are trained. A common training algorithm used for many applications is the backward error propagation algorithm, or simply the "backprop" algorithm (25:54). The training patterns are presented to the network along with the desired outputs from the MLP. The backprop algorithm updates the weights on the interconnections until either the error between the desired output d_j and the actual output calculated by the MLP y_j ,

$$E = \frac{1}{2} \sum_j (y_j - d_j)^2 \quad (45)$$

or until the classification error on an independent test set is reduced to a tolerable level.

Rogers *et al.* (25) give the following procedure for the backprop training algorithm:

1. Initialize the weights ω_{ij} and biases to small random numbers
2. Present inputs and desired outputs to the network
3. Calculate the output from the neural net by calculating Eqn. (43) and Eqn. (44) for each perceptron
4. Update the weights and biases using

$$\omega_{ij}^+ = \omega_{ij}^- + \eta \delta_j x_i + \alpha (\omega_{ij}^- - \omega_{ij}^{--}) \quad (46)$$

where ω_{ij}^+ is the updated weight, ω_{ij}^- is the previous weight, and ω_{ij}^{--} is the weight before the last update. α is the momentum, η is the learning rate, and x_i is the output of node i .

For a sigmoid function, δ_j is defined as (25:36)

$$\delta_j = \begin{cases} (d_j - y_j) y_j (1 - y_j), & \text{for output node } j \\ x_i (1 - x_i) \sum_k \delta_k \omega_{jk}, & \text{for hidden node } j \end{cases} \quad (47)$$

whereas for a hyperbolic tangent function

$$\delta_j = \begin{cases} (d_j - y_j) (1 - y_j^2), & \text{for output node } j \\ (1 - x_i^2) \sum_k \delta_k \omega_{jk}, & \text{for hidden node } j \end{cases} \quad (48)$$

where d_j is the desired output, y_j is the output calculated by the neural network, and x_i is the input to hidden node j .

After training the network by updating its weights, the MLP can be used as a classifier. Independent test samples are entered into the network and propagated through the interconnections and weights using Eqn. (43) and Eqn. (44) to yield the output from the MLP. The output signifies the class either directly or by some problem-specific classification rule.

3.3 Concepts for Constrained Linear Estimation

As discussed in the introduction to this chapter, the spectrum, or observation, may be modelled in matrix notation as

$$\vec{b} = A\vec{x} + \vec{n} \quad (49)$$

where \vec{x} is an M -dimensional column vector, \vec{b} and \vec{N} are L -dimensional column vectors, and A is an L -by- M matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ a_{L1} & a_{L2} & \cdots & a_{LM} \end{bmatrix} \quad (50)$$

with elements being the spectral reflectivities for the materials. For this research, the A matrix is 28-by-9 which yields an "overdetermined least-squares problem". For high SNR conditions, \vec{x} can be estimated in a least-squares sense from a simple matrix multiplication of the generalized inverse of the library matrix and the observed mixture

$$\hat{\vec{x}} = (A^t A)^{-1} A^t \vec{b} \quad (51)$$

where the main difficulties lie in computing the generalized inverse of A . If the elements of A are linearly dependent, or have small singular values, the estimate $\hat{\vec{x}}$ of \vec{x} will be inexact, and small changes in \vec{b} can cause large changes in $\hat{\vec{x}}$.

As noise levels increase, the estimate using Eqn. (51) becomes unreliable as can be shown by the expression for the root-mean-square (RMS) error

$$\epsilon = \frac{1}{2} (\hat{\vec{x}} - \vec{x})^t (\hat{\vec{x}} - \vec{x}) \quad (52)$$

Substituting the difference between the estimated and actual compositions

$$\hat{\vec{x}} - \vec{x} = (A^t A)^{-1} A^t \vec{b} - (A^t A)^{-1} A^t (\vec{b} - \vec{N}) \quad (53)$$

into Eqn. (52) gives

$$\epsilon = \frac{1}{2} \left((A^t A)^{-1} A^t \vec{N} \right)^t \left((A^t A)^{-1} A^t \vec{N} \right) \quad (54)$$

as the RMS error. Thus for low singular values in A , $(A^t A)^{-1} A^t$ becomes large and the RMS error increases to unacceptable levels. Other methods of estimating \vec{x} under low SNR conditions may be required.

An alternative method is to solve the problem in a least-squares sense and minimize the residual

$$r = ||A\vec{x} - \vec{b}||^2 \quad (55)$$

subject to the constraints

$$x_i \geq 0 \quad i=1,\dots,M \quad (56)$$

and

$$\sum_{i=1}^M x_i = 1 \quad (57)$$

A built-in MATLAB procedure in the MATLAB Optimization Toolbox was used in this thesis to solve Eqn. (55), using an algorithm given by Lawson and Hanson (16:351).

3.4 *Summary*

This chapter introduced the architectures and concepts chosen in this research for estimation of material composition. Three architectures were under investigation: a MLP designed for direct estimation, a constrained least-squares linear estimator, and a hybrid system using a MLP classifier and a linear estimator. Section 3.2 introduced pattern recognition concepts, specifically those related to MLP and Parzen classifier design. Also included in Section 3.2 was a discussion on feature extraction and data analysis. The Fisher discriminant approach was chosen for feature extraction, and the PCA approach for data analysis. The constrained linear estimation technique was discussed in Section 3.3. The following two chapters provide results for both a single material problem, Chapter IV, and a two material problem, Chapter V.

IV. The Single Material Problem: Classification

This chapter provides classification results on “single material” data, where the term “single material” refers to the situation where each data sample, or observation, consists of the spectrum of a single material that has been corrupted by an additive Gaussian noise, which is uncorrelated between spectral bins. Two classifiers were used in the single material problem: a Parzen classifier and a Multi-Layer Perceptron (MLP). The objectives were to baseline the MLP algorithm by comparing its performance with the Parzen classifier, and to determine the expected level of performance of the classifier at low SNRs.

The simulated data resembles the data that is obtained by an ideal Sagnac interferometer with the inclusion of an additive zero-mean Gaussian white noise and was generated as described in Section 4.1. Once the data was created, the features were extracted and analyzed as explained in Section 4.2. Section 4.3 furthers the discussion given in Section 3.2.2 on the Parzen classifier and provides results when applying this classifier to single material data. Section 4.4 provides a discussion on the design of the Multi-Layer Perceptron and gives the results when the network was tasked to identify the material as in Section 4.3 for the Parzen classifier. Finally, Section 4.5 provides a comparison of the results of the two classifiers and a brief summary.

4.1 Data Simulation

The data used throughout this research was simulated using a material database containing spectral reflectivities of nine common satellite materials (14). The spectral reflectiv-

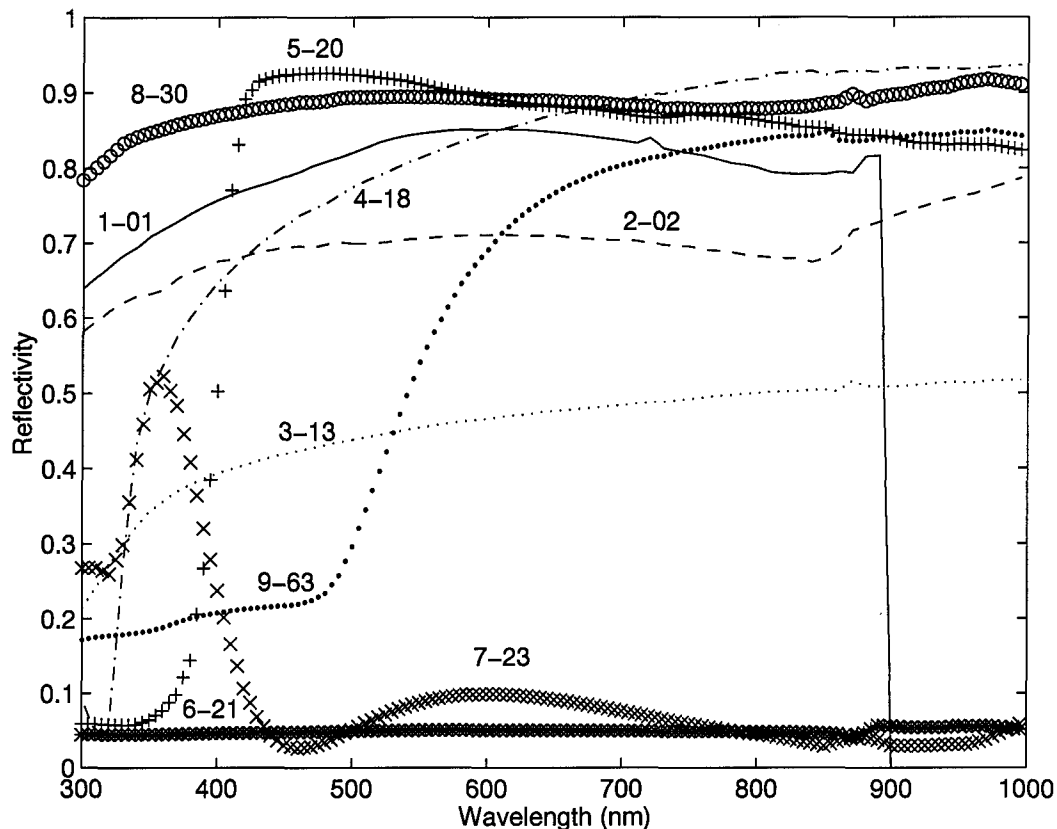


Figure 11. Spectral Reflectivities of the Nine Materials in the Material Database. The numbers that are plotted along with the data are the material designations (1 through 9) used throughout this thesis and their corresponding SMC numbers that were given in the original database.

ities of the materials were given in 5 nm increments for wavelengths between 300 nm and 1 μm , for a total of 140 spectral bands in each spectrum, see Fig. 11. The spectral reflectivities for each material were entered into the data simulation software in the form of vectors, where an element in the vector represented the spectral reflectivity at a particular wavelength of light incident on the Sagnac interferometer's input aperture. The entire material database

is represented by the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ a_{L1} & a_{L2} & \cdots & a_{LM} \end{bmatrix} \quad (58)$$

where the column vectors represent the spectral reflectivities for each of the M (i.e. 9) satellite materials, and the row vectors represent each of the L (i.e. 140) sampled wavelengths.

As discussed in the introductory section to Chapter 3, it was assumed that the photon interaction with the surface of the satellite behaves in a linear fashion. The data for the single material problem was therefore simulated using a linear relationship,

$$s_l = a_{ml} + \gamma n_l \quad l=1, \dots, L \quad (59)$$

which provides a simulated spectrum $\vec{s} = (s_1, \dots, s_L)^t$ consisting of the spectral reflectivity for the material \vec{a}_m corrupted by a zero-mean, unity-variance additive Gaussian noise, n_l . The noise term was selected randomly from a normal Gaussian distribution and γ is a constant that provides the variance in the data and allows the data to be simulated with the desired SNR. The SNR at a particular wavelength, or spectral bin i , is expressed as

$$SNR_i = \frac{\bar{s}_i}{\sigma_{s_i}} \quad (60)$$

Table 4. Average SNRs for Single Material Data. The SNRs are the ratio of the mean to the standard deviation over all samples in the data file. The Train and Test columns refer to the training data file and the test data file, respectively.

Material	$SNR = 1$		$SNR = 10$		$SNR = 100$	
	Train	Test	Train	Test	Train	Test
1	1.03	1.02	10.30	10.23	102.60	102.33
2	1.05	1.04	10.46	10.50	104.81	104.71
3	1.00	1.00	9.91	10.00	100.03	99.89
4	1.03	1.03	10.34	10.28	103.33	102.59
5	0.99	0.99	9.89	9.88	98.82	99.50
6	0.99	0.99	9.81	9.90	98.22	99.50
7	1.02	1.03	10.30	10.26	102.87	101.83
8	1.02	1.02	10.20	10.19	102.00	101.75
9	1.02	1.02	10.20	10.18	101.77	101.27

where \bar{s}_i is the mean and σ_{s_i} the standard deviation of the data set. In this research, \bar{s}_i was the spectral reflectivity in spectral bin i for one of the materials in the material database. γ was determined by trial-and-error so that the average spectral SNR of all samples within the data file was properly modeled. After γ was chosen such that the data modeled the appropriate SNR, the data was stored in a data file.

Each data file contained 4500 observations, where 500 observations were associated with each of the nine materials. The noise variance was chosen for data files with average SNRs of approximately 1, 10, and 100, as shown in Table 4. There were a total of six data files for the one material problem: a training file and a test file for each of the three SNR conditions. The SNR conditions of 1,10, and 100, will hereafter be referred to as the Low, Mid, and High SNR cases.

4.2 *Data Analysis and Feature Extraction*

As mentioned in the previous section, each data file contained 4500 observations and each observation contained 140 elements. The data contains the necessary information for discrimination, but the number of features in each sample must be reduced in order to decrease the complexity in the classifier stage, and to increase the practicality in obtaining the required number of training samples in an actual system. This section explains the feature extraction methods that were used to reduce the number of features from 140 to a more manageable number, 28, and thereby reduce the complexity of the problem. As discussed in Section 3.2, both principal components and Fisher discriminants were used for analysis. The following provides an analytical discussion on both techniques for the one material problem.

4.2.1 Principal Component Analysis. The data stored in each of the six data files had multimodal distributions. Each data sample belonged to a 500 sample Gaussian distribution, where the mean location of the distribution coincided with the entry for that material in the material database. Rather than calculating the principal components for the data in each data file, the principal components were calculated on the data in the material database. The eigenvalues for the 10 principal components are shown in Table 5. The magnitude of the eigenvalues correspond to the amount of variance present in the data along the direction of its eigenvector. To help visualize classification difficulty on a given data set, the data is projected into the plane containing the largest variance— the plane of

Table 5. Principal Eigenvalues for Material Database. The variances are ranked from largest to smallest magnitudes. The cumulative percent refers to percentage of total data variance found in eigenvalues up to current entry. All 140 eigenvalues summed to a value of 16.5753. The magnitude of each eigenvalue is rounded to four significant digits beyond the decimal point.

Rank	Magnitude	Cumulative %
1	13.1025	79.05
2	2.1268	91.88
3	0.7943	96.67
4	0.4584	99.44
5	0.0843	99.95
6	0.0081	100.00
7	0.0008	100.00
8	0.0003	100.00
9	0.0000	100.00
10	0.0000	100.00

the first two principal components. As shown in Table 5, this plane contains 91.88% of the total variation in the data.

Fig. 12 shows the projection of the data for the low SNR case. Only 20 samples are displayed for each of the nine materials. In this projection, the data points for materials 6 and 7 are clearly separated from the data for the other materials and pose little problem for the classifier, as will be shown later. However, the distributions for the other materials are overlapping and may lead to large classification errors in this data.

The same projection for data of mid SNRs is given in Fig. 13. The data for each material is more concentrated around its mean than for the low SNR case, as is expected for Gaussian distributions. In this figure, there are only two overlapping distributions, those of material 4 and material 5. Therefore, one can expect negligible classification errors for all

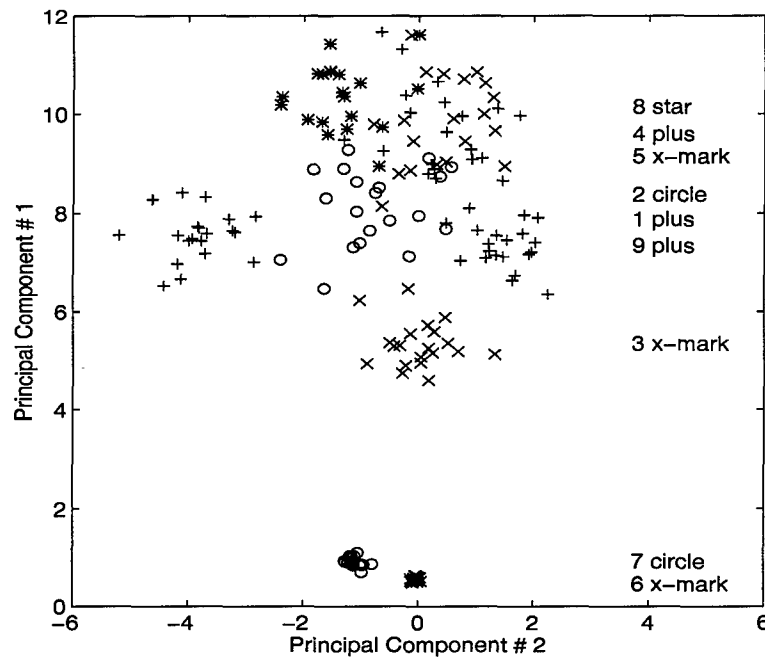


Figure 12. Single Material Training Data Projected into Plane of First Two Principal Components; Low SNR Case. The legend refers to the corresponding material designation and point type for data shown at same approximate height. Twenty observations are shown for each material.

class-pairs except the 4-5 pair. It is probable that the 4-5 pair will result in classification errors due to ambiguity caused by the overlapping distributions.

The projections for the high SNR data are shown in Fig. 14. Again, the data is more concentrated around its mean, and the distributions for all class-pairs are clearly separated. Therefore, on this data one would expect little or no classification errors for a properly designed classifier.

4.2.2 Fisher Discriminants. As an additional approach in data analysis, a generalized Fisher discriminant was calculated for each class comparison. The Fisher discriminant was used as a metric in determining the separation in the data, and also aided in the re-

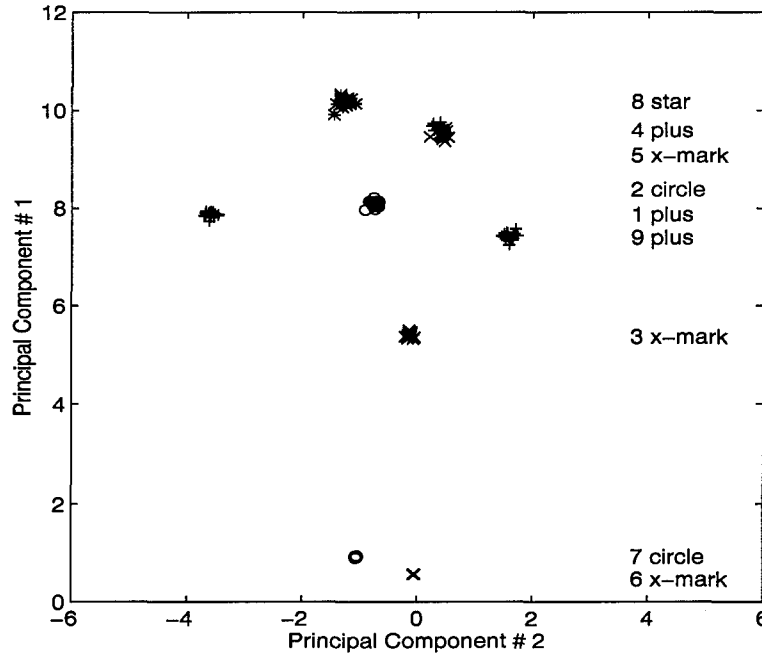


Figure 13. Single Material Training Data Projected into Plane of First Two Principal Components; Mid SNR Case. The legend refers to the corresponding material designation and point type for data shown at same approximate height. Twenty observations are shown for each material.

duction in the number of features presented to the classifier. As an example of the Fisher discriminant, consider one of the data files: it contains 4500 samples, 500 from each of the nine classes, designated class 1 through class 9. To compute the Fisher discriminant for data in class 1 against data in class 2, calculate

$$F = \sum_{i=1}^N f_{12i} \quad (61)$$

where

$$f_{12i} = \frac{(m_{1i} - m_{2i})^2}{\sigma_{1i}^2 + \sigma_{2i}^2} \quad (62)$$

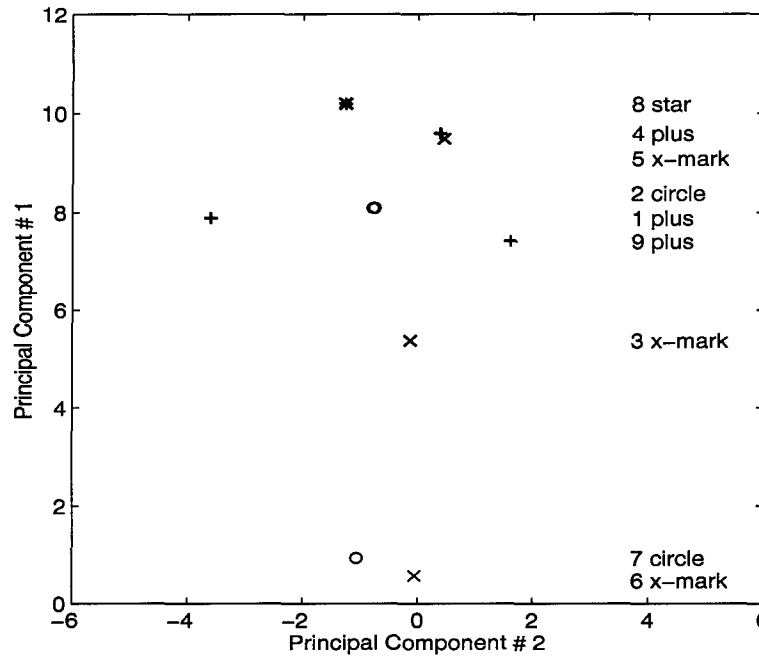


Figure 14. Single Material Training Data Projected into Plane of First Two Principal Components; High SNR Case. The legend refers to the corresponding material designation and point type for data shown at same approximate height. Twenty observations are shown for each material.

is the Fisher ratio corresponding to the class 1 and class 2 comparison for the i -th feature out of a total of N features. Since the data file contained samples for nine classes, 36 discriminants were obtained.

A generalized Fisher discriminant was also calculated for each of the 140 features using the equation

$$f^i = \sum_{j=1}^c f_{ji} \quad (63)$$

where f_{ji} was obtained for the i -th feature as in Eqn. (62) and c was the total number of class comparisons, 36. After the generalized discriminants were calculated, the 28 largest were used as inputs to the classifier.

The Fisher discriminants, both before and after feature extraction, calculated on the low SNR training data are shown in Table 6. Conceptually, the larger the magnitude of the discriminant, the greater the ability of any given classifier to correctly discriminate between the classes. Since classifier discrimination ability on data subsets are always worse than on the data itself, an increase in one of the entries in the table was not expected. Ideally, the number of features are reduced considerably while the magnitude of the generalized Fisher discriminant remains approximately the same. Shown in the lower portion of Table 6 are the magnitudes for the generalized Fisher discriminants of the 28 features providing the largest discriminants for this data set. An attempt to increase the magnitudes of the smaller entries in Table 6 by hand-picking features resulted in a reduction in overall classification accuracy for the system. Thus, the 28 features having the largest Fisher discriminants were used as inputs to the classifier. A comparison of the Fisher discriminants, all 140 features, for the low SNR case against those for the mid SNR case, Table 15, and the high SNR case, Table 16 (both located in Appendix A.1), illustrate that the classes with greatest similarity are class 4 and class 5. This result was also determined previously using PCA. However, after the features were reduced to 28, several class-pairs appear less separable than the 4-5 pair, including pairs 5-9, 4-8, 2-8, and 2-4. The significance of these pair combinations will become clear in the next section.

4.3 Performance of the Classifiers

The training data entered into the classifier consisted of 4500 samples of 28 features each. The features were chosen using Fisher discriminants explained in the previous section.

Table 6. Fisher Discriminants for Single Material Data; Low SNR Case

Discriminants using 140 Spectral Components									
Class	Class								
	1	2	3	4	5	6	7	8	9
1	0	16.04	32.61	21.36	22.44	153.58	130.31	15.32	37.42
2		0	13.81	5.32	9.56	133.39	114.67	4.30	13.47
3			0	23.63	26.21	113.16	97.15	27.88	18.42
4				0	3.45	139.50	129.28	4.65	10.15
5					0	138.83	135.97	9.18	13.95
6						0	184.83	129.78	152.17
7							0	113.86	146.49
8								0	19.83
9									0

Discriminants using 28 Spectral Components									
Class	Class								
	1	2	3	4	5	6	7	8	9
1	0	12.68	9.63	16.46	16.48	11.11	3.97	13.12	19.73
2		0	3.33	0.82	3.00	30.12	24.60	0.68	2.39
3			0	5.31	4.00	23.43	21.19	5.71	4.64
4				0	1.71	30.53	28.02	0.73	1.25
5					0	21.95	25.06	4.01	0.22
6						0	113.37	26.77	37.08
7							0	21.65	39.17
8								0	3.43
9									0

This section provides classification results on this data using two classifiers: a Parzen classifier, and a Multi-Layer Perceptron. The theory of the classifiers is presented in Chapter III. Presented in the following are issues regarding the solution to the single material problem, and the results that were obtained.

4.3.1 Results for the Parzen Classifier. As introduced in Section 3.2.2, the Parzen classifier models the class distributions in probabilistic terms. The single material problem consists of nine class distributions. The goal of the classifier is to identify the class that a sample, whose class origin is unknown, is a member of. Errors frequently occur in class associations of this sort, and a conventional means of displaying the errors are as entries in a confusion matrix. The row of the matrix associates with the actual class of the object, and the column of the matrix associates with the class assigned to the object by the classifier. Therefore, diagonal entries are correct classifications and off-diagonal entries are incorrect classifications.

The confusion matrix obtained by the Parzen classifier for low SNR data is shown in Table 7. The errors, or off-diagonal entries, are largely attributed to the pairs 5-9, 4-8, 2-8, and 2-4. As an example of the significance of this fact, of the possible 36 off-diagonal pairs, 46.7% of the 1655 errors are caused by the four pairs, or eight entries, that were flagged as being troubled pairs using Fisher discriminants and principal component basis vectors. This result illustrates the usefulness of these two approaches in increasing understanding of the data in the classifier design stage. In summary of the performance of the Parzen classifier,

Table 7. Confusion Matrix for Single Material Problem obtained from Parzen Classifier; Low SNR Case

Actual Class	Desired Class								
	1	2	3	4	5	6	7	8	9
1	437	6	54	0	2	0	0	0	1
2	3	221	79	35	34	0	0	7	121
3	15	2	475	0	0	0	0	0	8
4	1	140	8	112	75	0	0	31	133
5	0	48	44	35	110	0	0	3	260
6	0	0	0	0	0	500	0	0	0
7	1	0	0	0	0	0	500	0	0
8	1	166	5	112	38	0	0	149	29
9	1	38	95	2	22	0	0	1	341

the overall classification accuracies were 63.22% for low SNR data, 99.07% for mid SNR data, and 100.00% for high SNR data.

4.3.2 Results for the Multi-Layer Perceptron. The MLP was trained using the backprop training algorithm given in Chapter III. The architecture for the MLP consisted of 28 input nodes, 12 hidden nodes, and 9 output nodes, as illustrated in Fig. 15. The input nodes and hidden nodes both used the sigmoid function for nonlinearity. The approximate upper limit on the number of hidden nodes was determined to be 11.6 based upon the rule proposed by Baum: the number of weights in the network should be less than one-tenth the number of samples (1:153). The actual number of hidden nodes used was determined by trial-and-error by choosing the network configuration which provided the lowest classification errors. A separate neural network was trained for each of the three SNR cases.

During the training phase, the samples were entered into the network in random order, and the perceptrons transformed the weighted sums of the inputs using a sigmoid function.

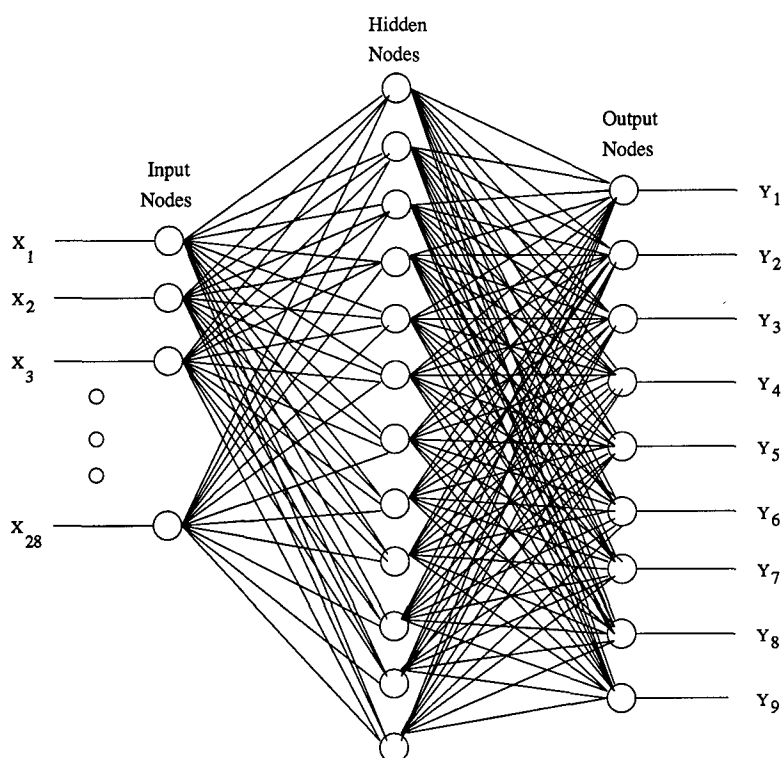


Figure 15. MLP Architecture for Single Material Problem

The weights in the network were updated based upon the error between the actual and desired outputs. The desired output of a single output node was activated and the desired output of the remaining eight nodes were unactivated. The activated node corresponded to the class the sample originated from. The neural nets were trained for 1000 complete passes, or epochs, through the training set.

The confusion matrix obtained by the MLP for low SNR data is shown in Table 8. As with the Parzen classifier, the off-diagonal entries are largely attributable to the pairs 5-9, 4-8, 2-8, and 2-4: 55.1% of the 1528 errors are caused by these four pairs. The overall classification accuracies are 78.04% for the low SNR case, 99.51% for the mid SNR case, and 100.00% for the high SNR case.

Table 8. Confusion Matrix for Single Material MLP Classifier on Data with Average SNR of One.

Actual Class	Desired Class								
	1	2	3	4	5	6	7	8	9
1	457	3	8	0	1	3	26	2	0
2	4	240	41	51	15	0	0	99	50
3	11	47	390	0	16	0	0	3	33
4	1	101	13	147	56	0	0	118	64
5	2	38	35	35	233	1	0	14	142
6	0	0	0	0	0	500	0	0	0
7	1	0	0	0	0	0	499	0	0
8	5	101	9	94	14	0	0	265	12
9	0	53	18	44	136	0	0	8	241

Table 9. Classification Accuracies for Single Material Problem. Values shown are percentage correct classification on a test set of 4500 samples.

<i>SNR</i>	<i>Parzen</i>	<i>MLP</i>
<i>Low</i>	63.22	78.04
<i>Mid</i>	99.07	99.51
<i>High</i>	100.00	100.00

4.3.3 *Comparisons of Results for the Two Classifiers.* The classification accuracies for the Parzen classifier and the MLP are shown side-by-side in Table 9. The performance of the two classifiers on the mid and high SNR data files are very agreeable. The difference in the results for the low SNR case can be attributed to three primary causes: the increased noise in the data, systematic errors in the training methods of the two classifiers, and the fact that the results for the MLP represent one sample of a stochastic random process. The results of the Parzen classifier for the low SNR data was very sensitive to the width of the Parzen window. A small change in window width resulted in a very large change in classification accuracies. Nonetheless, the comparable results in Table 9, especially for the

mid and high SNR cases, indicate that the MLP performs as expected.

4.4 Summary

Included in this chapter were the single material classification results for the Parzen classifier and the Multi-Layer Perceptron. Prior to the presentation of the results, an in-depth examination of data simulation, and feature extraction, was presented. The analysis using Fisher discriminants and principal components gave valuable insight into potential system classification errors. For the low SNR data, both approaches identified the four class-pairs providing the largest classification errors. However, overall the classification performance of the Parzen and the MLP agreed favorably, especially for the mid and high SNR cases. The approach taken in analyzing the single material data will now be applied to the percent composition estimates using two material data.

V. The Two Material Problem: Abundance Estimation

In this chapter, the observation contains the spectra of two of the nine materials. A re-cap of the three architectures considered for this problem is discussed in Section 5.1. The MLP approach is given in Section 5.2, and data analysis and feature extraction for the remaining two approaches are addressed in Section 5.3. The results of the linear estimator approach are given in Section 5.4, and Section 5.5 provides the results for the combined system. Finally, Section 5.6 yields a comparison of results and a brief summary.

5.1 The Three Architectures Revisited

The solution to the two-material problem requires knowledge of the materials that contributed to the observation as well as an estimate of their compositions, x_1 and $x_2 = 1 - x_1$. This chapter analyzes the two-material problem by implementing the systems proposed in Section 3.1. The MLP approach uses an artificial neural network designed such that the desired output equals the percent composition. Therefore, the actual output from the network is an estimate of the material composition. The linear estimation approach uses a constrained least-squares linear approach for estimation. The hybrid approach consists of a MLP classification stage followed by a constrained least-squares estimator. The remaining sections in this chapter discuss the results for system implementation.

5.2 MLP Approach

As stated in the previous section, the MLP approach uses an artificial neural net to directly estimate the percent composition. The samples introduced to the network contained a linear combination of the spectra for materials 1 and 2. The composition of material 1 was determined through a random number generator, and that of material 2 was constrained such that the total compositions for the two materials summed to one. Therefore, the two material data is described by the expression

$$s_i = x_1 a_{1i} + (1 - x_1) a_{2i} + \gamma n_i \quad i=1, \dots, N \quad (64)$$

where $\vec{s} = (s_1, \dots, s_N)^t$ represents the observation, x_1 represents the composition for the first material, and \vec{a}_1 and \vec{a}_2 are the elements of the material spectral database for the two materials contributing to the mixture, respectively, as given in Eqn. (59). γ and n_i are also as defined for Eqn. (59).

Eqn. (64) was used to create 5000 observations consisting of random contributions from material 1 and material 2. Rather than using the entire 140 spectral bins in simulation, every fifth bin was used, which yielded data consisting of 28 dimensions. All 28 dimensions were entered into the neural network as features. The interconnection diagram for the MLP is shown in Fig. 16. Two outputs were used, each representing the estimate of the percent composition for one of the two materials. The backprop training algorithm performed the weight update where the perceptrons used the sigmoid function for nonlinearity. The MLP was trained for 1000 epochs. The results applied to the test data, shown in Fig. 17, indicate

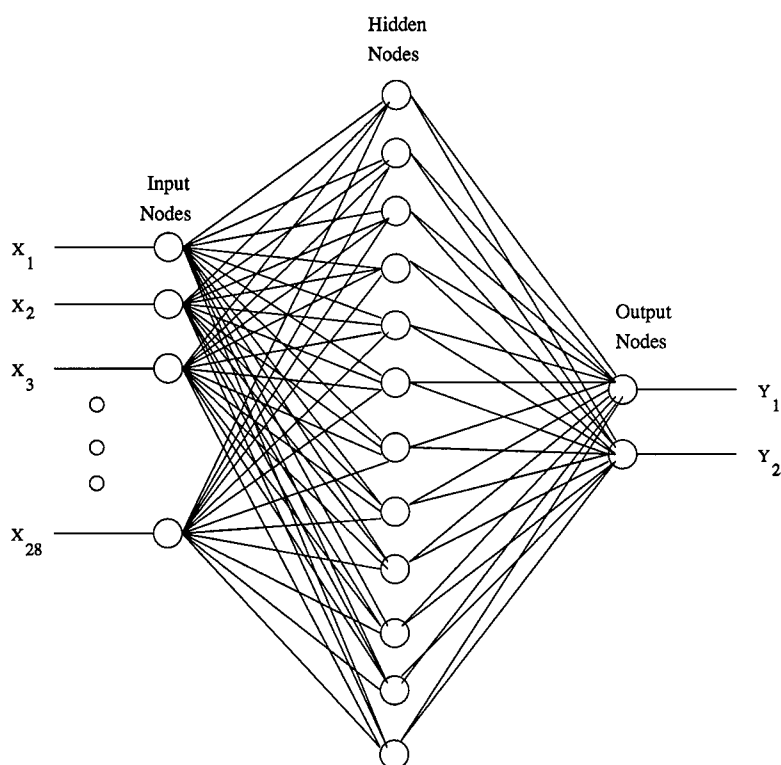


Figure 16. MLP Architecture for Two Material Problem.

the level of performance of the MLP for estimation of percent composition. The estimate at low SNRs is corrupted by the randomness of the data, as is clearly shown.

Due to the poor results of the MLP approach and the time required for training the weights, the MLP was eliminated from consideration as a practical device for direct estimation. The number of required training samples for problems dealing with concentrations of more than two materials was simply overwhelming. Also, the poor performance of the MLP in a very controlled environment, as presented here, yields little hope of practical implementation of the MLP in this application. The following section discusses the data simulation, analysis, and feature extraction for the remaining two approaches used in this research.

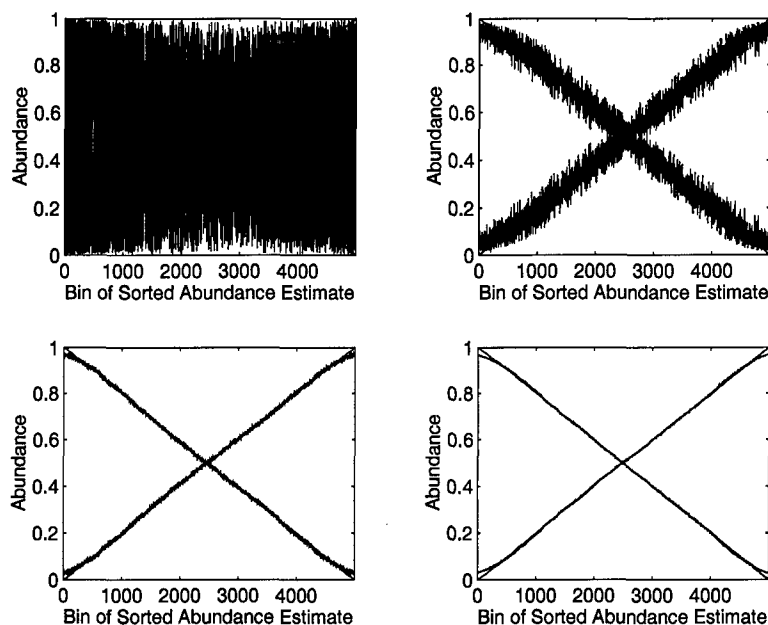


Figure 17. Estimation of Percent Composition using the MLP. The straight diagonal lines across each subplot represent the actual percent composition, or desired outputs for the MLP, and overlaid on the diagonals are the compositions estimated by the MLP. The subplots represent data with the following average SNRs: a) 1, b) 10, c) 100, and d) 1000.

5.3 Data Simulation, Analysis, and Feature Extraction

This section provides analysis on the two material data similar to that performed on the single material data in Chapter IV. Included in this section are data simulation, data analysis using principal components, and feature extraction using Fisher discriminants.

5.3.1 Data Simulation. The data was simulated using Eqn. (64). Each data file contained 4500 samples. There were a total of 36 possible combinations of the nine materials taken two at a time, and therefore each two-material grouping, or set, had 125 samples. The generation of large data files with several requirements, such as random concentrations, 36 subsets, and constant SNRs, becomes very computationally intense. Therefore, rather than

forcing the data file to have a constant SNR level, the samples within the data file were generated using the single material data. Two samples from the single material data were scaled by the appropriate percent composition and summed to yield the composite spectrum. As an example, let \vec{b}_i represent an observation of single material data for material i , and \vec{b}_j represent an observation of single material data for material j . A two material observation containing both the spectrum of material i and the spectrum of material j is modeled as

$$\vec{b} = x_1 \vec{b}_i + x_2 \vec{b}_j \quad (65)$$

where x_1 is the abundance for material 1, and $x_2 = 1 - x_1$ is the abundance for material 2. Both \vec{b}_i and \vec{b}_j contain an additive zero-mean Gaussian noise term, which results in a variable spectral SNR for the data. The spectral SNR lies in the range $1 < SNR < 1.4$ for the low SNR data, the range $10 < SNR < 14$ for the mid SNR data, and the range $100 < SNR < 140$ for the high SNR data. However, due to additional variance introduced in each data file by random material concentrations, the effective SNR of these files are much lower.

5.3.2 Principal Component Analysis. The two material data has characteristics that are desirable for principal component analysis. When the endmembers, defined as the spectra for the materials in the material database, are projected into the principal plane, a scatter plot results as shown in Fig. 18. Under noise-free conditions, a composite of two materials will yield a data point located at a point between the two corresponding endmembers. The ratio of distances between the data point and the endmembers is directly

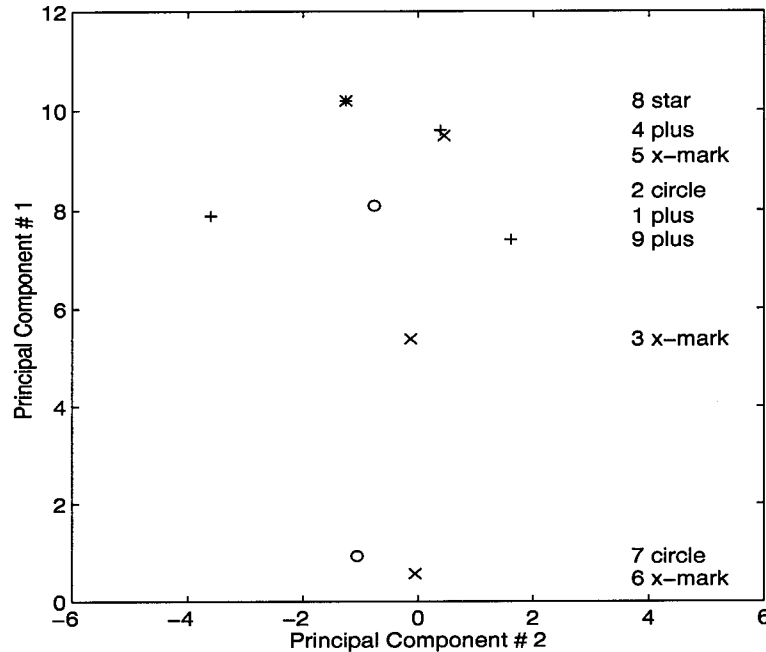


Figure 18. Endmembers Projected into PC Plane

related to the ratio of the compositions attributed to the endmembers. For example, a composite consisting of 50% material 1 and 50% material 7 will be located at coordinate $(-2.54, 6.96)$, midway between the two endmembers. With the addition of a zero mean Gaussian noise, the distribution of all similar data points will be located in a cloud of points centered around their mean location and whose variation is dependent on the magnitude of the noise variance.

The two material training data is shown projected in the principal component plane in Figs. 19-21. The pluses represent projections for material 1 data, and the minuses represent projections for data of other materials. The projections for the low SNR data, Fig. 19, are distributed in random fashion rather than linearly as discussed above. However, the projections for the mid and high SNR cases, Figs. 20 and 21, do show the linear characteristic

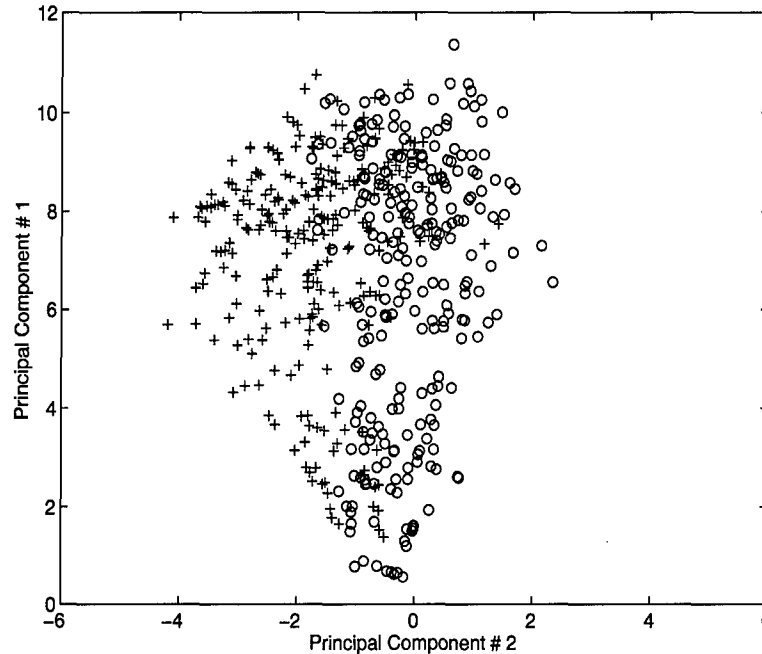


Figure 19. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 1 data shown with pluses and other material data shown with circles.

as described in the previous paragraph. Analysis on the scatter plots yield a visual indication of the classification difficulty for the two material problem. The projected data for material 1 have overlapping distributions. A survey of similar scatter plots for materials 2 through 9, Figs. 25–48, give similar conclusions. In the areas of overlap, a well-trained classifier will continue to misclassify samples. Because of this apparent overlap in the principal component plane, classification errors will tend to be high on this data. In addition, the low accuracies for the low SNR case will also result because the distributions in the principal component plane are more random and it is more difficult to fit a linear decision boundary between the different classes.

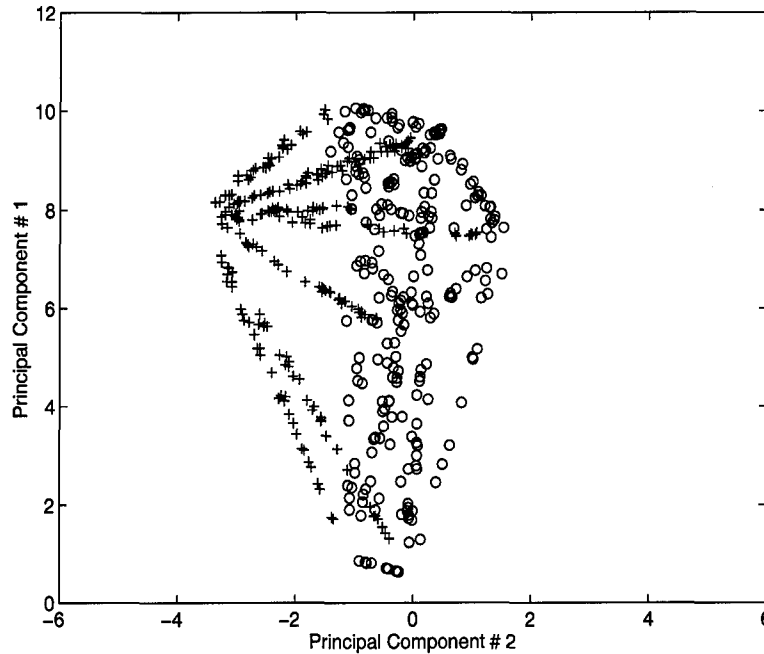


Figure 20. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 1 data shown with pluses and other material data shown with circles.

5.3.3 Fisher Discriminants. The Fisher discriminants were calculated for the two material data as given by Eqn. (63). After the Fisher discriminants for each of the 140 features in a data file were calculated, their sum was regarded as a generalized discriminant for the file. As was performed for the one material data, the features were then extracted by choosing the 28 features that provided the largest discriminants. A comparison of generalized Fisher discriminants for each data file gives an indication of the relative level of difficulty in designing a classifier for the data within each. The generalized Fisher discriminants both before and after feature extraction are given in Table 10. A comparison of the values in Table 10 to those of the single material data, Tables 6, 15, and 16, show that the classification performances at low SNRs are expected to be comparable, but at mid and high SNRs,

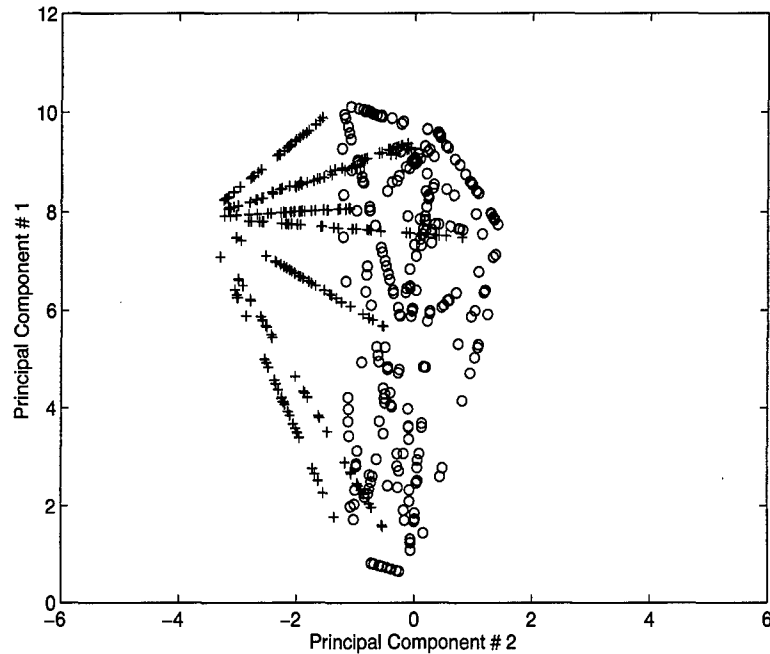


Figure 21. Two Material Training Data Projected into PC Plane: High SNR Case. Material 1 data shown with pluses and other material data shown with circles.

the performance for the two material data should be much lower than that for the single material data. The reason for this is that each sample now contains a linear combination of two material spectra. The distributions for each material in the two material composites are no longer concentrated in a small region in feature space, as is clearly illustrated in the analysis of the previous section. This effect is lessened on the low SNR data because of the inherent randomness in this data.

Also shown in Table 10 are indications that the largest classification errors should associate with classifications on material 2 and material 3, and the lowest errors should associate with material 6 and material 7.

Table 10. Generalized Fisher Discriminants for Two Material Problem

Material	140 Features			28 Features		
	<i>Low</i>	<i>Mid</i>	<i>High</i>	<i>Low</i>	<i>Mid</i>	<i>High</i>
1	6.27	33.86	36.60	3.40	16.16	16.85
2	2.01	9.81	10.13	1.05	5.12	5.39
3	1.06	4.29	4.97	0.44	1.46	1.69
4	5.35	28.22	29.55	2.05	8.94	9.45
5	6.62	37.20	38.60	2.31	13.77	14.14
6	25.68	149.37	141.94	6.97	38.96	36.80
7	22.95	104.07	124.20	6.89	30.47	36.28
8	9.03	59.65	67.34	3.14	26.19	29.84
9	4.53	24.75	24.93	1.67	8.97	8.94

5.4 Linear Estimator Approach

As discussed in Chapter III, a direct solution to the estimation problem is to solve it in a least-squares sense by minimizing the residual

$$r = \|A\vec{x} - \vec{b}\|^2 \quad (66)$$

subject to the constraints

$$x_i \geq 0 \quad i=1,\dots,M \quad (67)$$

and

$$\sum_{i=1}^M x_i = 1 \quad (68)$$

The results obtained by using this approach are given in Table 11. The mean errors for the low SNR data are 28.55% in estimating compositions in material presence and 13.23% in estimating compositions in material absence. The greatest RMS errors are in detecting

Table 11. Errors in Estimates using Linear Technique. ϵ_1 refers to the RMS error of the estimate when spectrum of material present in composite. ϵ_0 refers to RMS error of estimate when spectrum of material not present in composite. ϵ_t is the total RMS error for the estimate. The mean values are the average RMS error across all nine materials.

Material	Low SNR			Mid SNR			High SNR		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
1	.1313	.0833	.0961	.0153	.0095	.0111	.0016	.0009	.0011
2	.4396	.1892	.2661	.1996	.0756	.1153	.0225	.0088	.0132
3	.4995	.0636	.2421	.2675	.0167	.1270	.0371	.0022	.0176
4	.3513	.1476	.2106	.0538	.0325	.0382	.0062	.0042	.0047
5	.2230	.1265	.1533	.0293	.0187	.0215	.0029	.0020	.0023
6	.1746	.1020	.1219	.0469	.0551	.0534	.0057	.0078	.0074
7	.1962	.2097	.2068	.0340	.0316	.0321	.0032	.0034	.0033
8	.3824	.1277	.2125	.1084	.0716	.0812	.0127	.0080	.0092
9	.1718	.1414	.1487	.0210	.0196	.0199	.0022	.0020	.0021
mean	.2855	.1323	.1842	.0862	.0368	.0555	.0105	.0044	.0068

presence of either material 2 or material 3, and are approaching 50% for the low SNR case. As previously stated, the generalized Fisher discriminants given in Table 10 are also lowest in value for the material 2 and material 3 entries. Surprisingly, the entries in Table 11 for material 1 and material 9 are comparable to those of materials 6 and 7, a result not apparent from Fisher discriminants alone. Overall, the constrained linear estimator provided estimates with average errors of 18.42% for the low SNR case, 5.55% for the mid SNR case, and 0.68% for the high SNR case.

5.5 Hybrid Approach

The hybrid approach to estimation of percent composition is discussed in this section, with reference to the configuration illustrated in Fig. 22. The classification errors for the

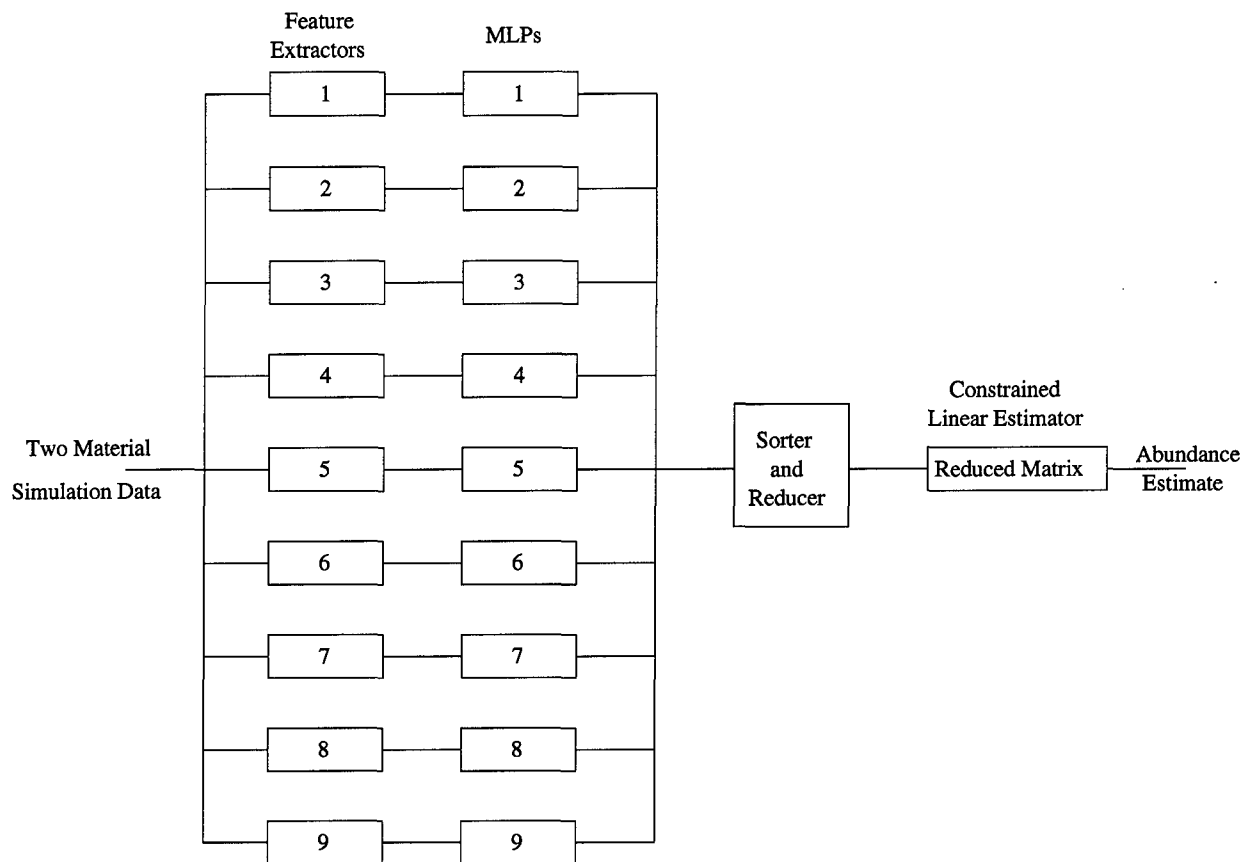


Figure 22. Block Diagram for Hybrid Approach to Estimation of Percent Composition.

MLP are given in Table 12. Each two element entry in Table 12 allows the complete reconstruction of the confusion matrix obtained in classification. For instance, the first row of data has ϵ_1 and ϵ_0 values equal to .159 and .0063, respectively. ϵ_1 refers to the fraction of the 1000 samples, containing the scaled spectrum for material 1, that were misclassified. Hence, there were 159 misclassifications and 841 correct classifications when the spectrum of material 1 was included in the composite. ϵ_0 refers to the fraction of the remaining 3500 samples (those which do not contain scaled spectrum of material 1) that were misclassified. Therefore, 22 of the 3500 samples were mis-tagged, leaving 3478 that were properly identified

Table 12. Error Rates for Confusion Matrix Resulting from MLP Classifier. ϵ_1 refers to the rate at which the classifier determines material not present, when the material actually was present. ϵ_0 refers to rate at which classifier determines material was present, when the material actually was not present. An ϵ_1 of zero indicates that all 1000 samples containing the materials spectrum were properly classified, and an ϵ_0 of zero indicates that all 3500 samples not containing the materials spectrum were properly tagged as such.

No.	Training Data						Test Data					
	Low SNR		Mid SNR		High SNR		Low SNR		Mid SNR		High SNR	
	ϵ_1	ϵ_0	ϵ_1	ϵ_0	ϵ_1	ϵ_0	ϵ_1	ϵ_0	ϵ_1	ϵ_0	ϵ_1	ϵ_0
1	.159	.0063	.059	.0069	.010	.0006	.421	.0577	.179	.0271	.022	.0011
2	.361	.0229	.229	.0486	.073	.0274	.743	.1169	.390	.1031	.091	.0360
3	.424	.0169	.233	.0383	.201	.0426	.753	.1026	.404	.0980	.221	.0514
4	.416	.0371	.326	.0254	.047	.0034	.723	.1240	.447	.0654	.061	.0091
5	.254	.0249	.053	.0003	.000	.0000	.596	.1086	.071	.0074	.000	.0000
6	.260	.0386	.349	.0437	.149	.0191	.525	.1143	.445	.0717	.162	.0226
7	.367	.0414	.354	.0471	.225	.0177	.546	.1049	.466	.0734	.213	.0157
8	.229	.0283	.221	.0140	.128	.0026	.572	.1054	.343	.0417	.138	.0040
9	.194	.0103	.417	.0446	.167	.0183	.446	.0891	.624	.0811	.176	.0266

as not containing the spectrum of material 1 within the composite. The confusion matrix for this entry, low SNR training data for material 1, is therefore given as

$$cm = \begin{bmatrix} 841 & 159 \\ 22 & 3478 \end{bmatrix} \quad (69)$$

In comparison of all entries in Table 12, the largest entries correspond to those of ϵ_1 . This is a direct result of allowing the weights in the MLP to be updated equally by the magnitude of the squared errors for each sample entered into the network. Because there were three-and-a-half times as many samples not containing the spectrum of the material as opposed to those that did, the network was biased toward lower values of ϵ_0 , rather than

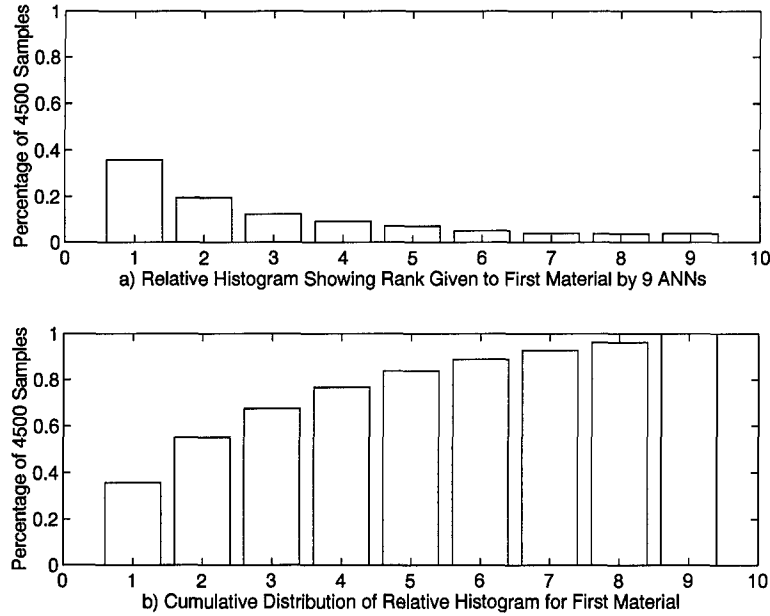


Figure 23. ANN Rankings for First Material in Two Material Composite for the Low SNR Case.

lower values of ϵ_1 . This is a classical probability of detection and probability of false alarm design trade. If estimation errors are mainly caused by large values of ϵ_1 , then more emphasis must be placed on lowering ϵ_1 by choosing an appropriate weight update strategy. However, lower ϵ_1 values generally imply larger ϵ_0 values, and the design requires choosing training strategies and comparing estimates obtained from each strategy. This comparison will yield the design providing the lowest errors. In this research, little time was devoted to this design trade.

A summary of the errors in abundance estimates for the hybrid approach is provided in Table 13 for the low SNR case, and Tables 17 and 18 for the mid and high SNR cases. The values for n represent the number of endmembers whose spectrum were allowed to be a component of the A matrix in Eqn. (66), or the follow-on linear estimation stage. The

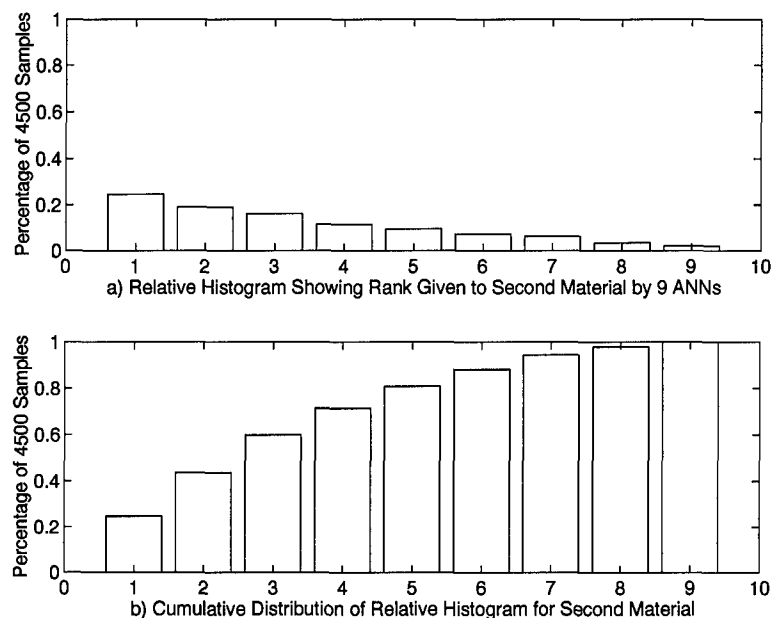


Figure 24. ANN Rankings for Second Material in Two Material Composite for the Low SNR Case.

actual n endmembers which contributed were determined by the output of the nine neural networks. Each neural net was designed to determine presence or non-presence of a certain material spectrum within the composite. The sample was introduced to all nine networks, and each network determined if its associated material spectrum was present. Outputs from the MLP approaching the value of one represented material presence, and output values approaching the value of zero represented material absence. The outputs from all nine networks were rank ordered and the materials associated with the largest n outputs were used in the A matrix. Figs. 23a and 24a show the relative importance, or ranking, assigned to the two materials whose spectra were actually present in the composite, for the low SNR case. Similar plots for the mid SNR and high SNR cases are given in Figs. 49–52. Fig. 23b and 24b illustrate that for the low SNR case, all nine endmembers must be included in the A

Table 13. Errors in Estimates using Hybrid Approach; Low SNR Case. n spectra were used in the database entered into the linear estimator. ϵ_1 refers to the RMS error of the estimate when spectrum of material present in composite. ϵ_0 refers to RMS error of estimate when spectrum of material not present in composite. ϵ_t is the total RMS error for the estimate. The mean values are the average RMS error across all nine materials.

Material	$n=3$			$n=4$			$n=5$		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
1	.2153	.0634	.1158	.1951	.0612	.1067	.1782	.0648	.1016
2	.4431	.2367	.2953	.4295	.2276	.2851	.4225	.2152	.2751
3	.3926	.2889	.3149	.3864	.2258	.2699	.4164	.1661	.2449
4	.3539	.2025	.2444	.3332	.1909	.2303	.3241	.1832	.2224
5	.3405	.1051	.1854	.3081	.1087	.1740	.2826	.1115	.1656
6	.2626	.1017	.1529	.2246	.0879	.1312	.1942	.0990	.1265
7	.2542	.1195	.1596	.2233	.1243	.1519	.2148	.1447	.1629
8	.4096	.1480	.2331	.4021	.1439	.2281	.3882	.1377	.2196
9	.2814	.1054	.1620	.2561	.0950	.1470	.2326	.1014	.1415
mean	.3281	.1524	.2070	.3065	.1406	.1916	.2948	.1360	.1845
Material	$n=6$			$n=7$			$n=8$		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
1	.1629	.0705	.0988	.1426	.0749	.0942	.1377	.0792	.0954
2	.4243	.2111	.2732	.4269	.2047	.2704	.4343	.1996	.2700
3	.4471	.1342	.2417	.4737	.1154	.2454	.4896	.0982	.2465
4	.3285	.1747	.2185	.3341	.1660	.2150	.3408	.1575	.2124
5	.2632	.1204	.1633	.2438	.1234	.1583	.2250	.1267	.1541
6	.1786	.1040	.1245	.1749	.1023	.1222	.1740	.1018	.1216
7	.2063	.1711	.1795	.2010	.1912	.1935	.1977	.2046	.2031
8	.3910	.1334	.2186	.3912	.1278	.2161	.3892	.1268	.2148
9	.2200	.1102	.1421	.2077	.1177	.1427	.1952	.1267	.1447
mean	.2913	.1366	.1845	.2885	.1359	.1842	.2870	.1357	.1847

matrix if misclassification errors are not to increase errors in the estimation of the abundance. Similarly, the mid and high SNR data require all nine endmembers, but the increased errors in abundance estimation attributed to misclassification should not be as noticeable for lower number of endmembers included in A because the cumulative distributions have almost completely converged for n equal to six.

The errors shown in Table 13, 17, and 18, all follow similar trends: from low to high values as n decreases from eight to three. Therefore, in this design, the misclassifications due to the MLP are directly degrading the performance of the linear estimator.

5.6 Summary

Table 14 is provided for additional comparisons of the linear estimator approach to the hybrid approach. An n equal to nine represents the linear approach, and the row for n equal to two illustrates the best possible performance of the linear estimator on this data. This row assumes that the linear estimator has a classifier stage preceding it that has 100% classification accuracy. Therefore, the A matrix can be reduced to two columns, each column associated with one of the two materials present within the composite.

The table clearly illustrates that the linear estimator provides lower RMS errors, as opposed to the Hybrid approach, for all three data files. The trend of increasing errors as n decreases is apparent in each column. This trend is a direct result of the large classification errors resulting from the MLP. Some of the errors in ϵ_1 , shown in Table 14, are due to the weight update method implemented for the MLP. Further investigation should address the

Table 14. Comparison of mean RMS Errors; Linear Approach vs Hybrid Approach. ϵ_1 refers to the mean RMS error of the estimate when spectrum of material present in composite. ϵ_0 refers to mean RMS error of estimate when spectrum of material not present in composite. ϵ_t is the total mean RMS error for the estimate. n is number endmembers included in database before linear estimation stage. An n equal to nine (i.e. top row), the complete database, uses the linear approach to abundance estimation.

n	Low SNR			Mid SNR			High SNR		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
9	.2855	.1323	.1842	.0862	.0368	.0555	.0105	.0044	.0068
8	.2870	.1357	.1847	.0910	.0371	.0572	.0246	.0129	.0202
7	.2885	.1359	.1842	.0926	.0372	.0577	.0260	.0140	.0217
6	.2913	.1366	.1845	.0966	.0383	.0590	.0339	.0178	.0241
5	.2948	.1360	.1845	.1103	.0461	.0671	.0429	.0244	.0309
4	.3065	.1406	.1916	.1424	.0600	.0868	.0577	.0279	.0387
3	.3281	.1524	.2070	.1855	.0847	.1168	.0784	.0316	.0477
2			.1884			.0282			.0031

weight update issue and attempt to lower the values for the ϵ_1 and ϵ_0 entries in Table 13 such that the decision made by the MLP aids the performance of the linear estimator and reduces the errors in Table 14 as n decreases.

VI. Conclusions and Recommendations for Future Research

The research conducted in this thesis had one overall objective: to determine if the interferogram obtained by the Sagnac interferometer could be used by a system to estimate satellite material compositions. The approach implemented first introduced the Sagnac interferometer and determined its expected level of performance. Based upon the estimated level of performance, data was simulated that had representative noise levels.

The simulated data was used in two problems: a single material problem, and a two material problem. The single material problem was developed to introduce pattern recognition concepts and to baseline the performance of the MLP. The two material problem yielded the percent composition estimates for three proposed architectures: a direct approach using the MLP, a constrained least squares approach, and a hybrid approach.

Of the three approaches, the constrained least squares approach provided estimates with the lowest RMS errors: 18.42% for low SNR data, 5.55% for mid SNR data, and 0.68% for high SNR data. Also, the constrained least squares approach is most practical in terms of ease of implementation. In this thesis, as developed and presented, the hybrid approach yielded slightly larger errors, 18.47% for low SNR data, 5.72% for mid SNR data, and 2.02% for high SNR data, than those obtained using the linear estimator alone. In the presence of non-ideal effects, such as atmospheric scintillation, and system transfer function, the hybrid approach may provide a more robust design that makes it a more practical approach.

With this in mind, several issues are left unanswered and may provide topics for further investigation:

- Will a new weight update strategy improve classifier performance such that the hybrid approach provides better estimates?
- Will the introduction of non-ideal circumstances into the model show that the hybrid approach provides better estimates?
- How does the approach presented in this thesis relate to actual data measured by the Sagnac interferometer?

The material in this thesis, as presented, clearly illustrate that the Sagnac interferometer can be used for material composition estimates under ideal circumstances. The percent errors are tolerable (i.e. $< 10\%$) for data with average SNRs of 10 or greater, and radiometric results, Table 3, indicate that these SNRs are possible with this instrument. The questions stated above will help determine whether the Sagnac interferometer can be used for estimation of material compositions under more practical circumstances.

Appendix A. Analysis and Results

A.1 Fisher Discriminants for Single Material Data

Table 15. Fisher Discriminants for Single Material Data; Mid SNR Case

Discriminants using 140 Spectral Components									
Class	Class								
	1	2	3	4	5	6	7	8	9
1	0	1509	3172	2039	2106	15308	12916	1463	3684
2		0	1311	507	880	13213	11374	402	1298
3			0	2327	2517	11073	9497	2717	1751
4				0	298	14030	13003	420	992
5					0	13680	13361	836	1373
6						0	18444	12948	15224
7							0	11405	14649
8								0	1915
9									0

Discriminants using 28 Spectral Components									
Class	Class								
	1	2	3	4	5	6	7	8	9
1	0	1324	969	1720	1605	881	223	1380	2081
2		0	280	71	216	2915	2480	65	197
3			0	488	327	2390	2213	511	450
4				0	139	3026	2813	55	108
5					0	2144	2398	309	9
6						0	10280	2595	3870
7							0	2190	4083
8								0	281
9									0

Table 16. Fisher Discriminants for Single Material Data; High SNR Case

Discriminants using 140 Spectral Components									
Class	Class								
	1	2	3	4	5	6	7	8	9
1	0	149900	316600	204800	210400	1521200	1285300	148300	365400
2		0	131600	51600	89100	1325700	1140100	40400	129400
3			0	233500	253100	1127200	966300	272700	175600
4				0	29800	1400800	1299000	43000	98700
5					0	1366300	1334600	84800	135000
6						0	1866500	1295800	1514800
7							0	1138600	1457400
8								0	190800
9									0

Discriminants using 28 Spectral Components									
Class	Class								
	1	2	3	4	5	6	7	8	9
1	0	131300	97300	172100	160200	84600	21200	140000	207600
2		0	28100	7300	22700	287400	241800	6600	20600
3			0	49700	33000	244700	227000	52300	45700
4				0	14000	305500	284700	5500	10800
5					0	209700	235800	32000	900
6						0	1043000	264000	385300
7							0	221600	407500
8								0	29100
9									0

A.2 Two Material Data shown Projected in Two Principal Component Plane

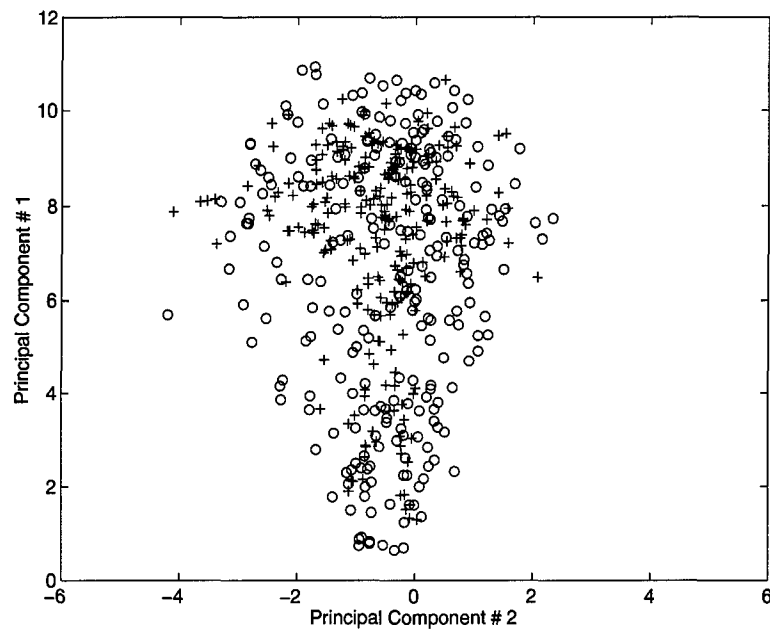


Figure 25. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 2 data shown with pluses and other material data shown with circles.

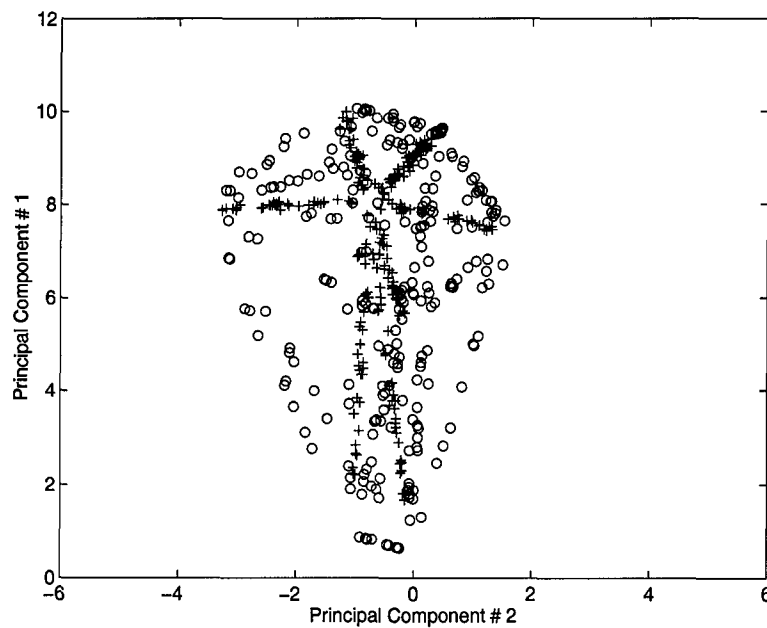


Figure 26. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 2 data shown with pluses and other material data shown with circles.

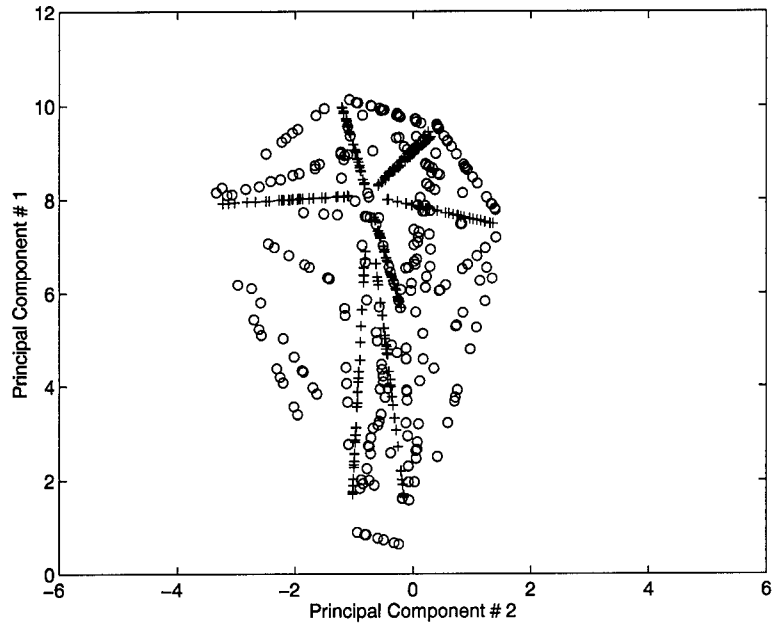


Figure 27. Two Material Training Data Projected into PC Plane: High SNR Case. Material 2 data shown with pluses and other material data shown with circles.

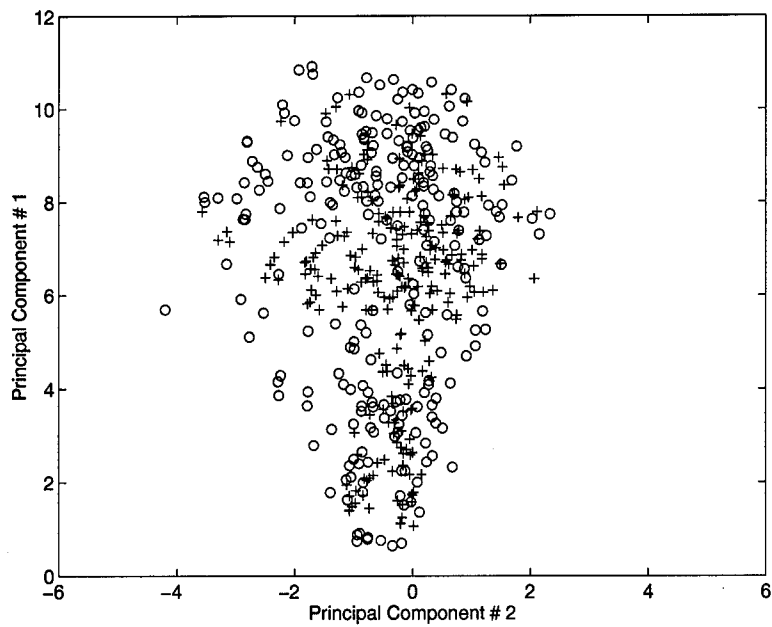


Figure 28. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 3 data shown with pluses and other material data shown with circles.

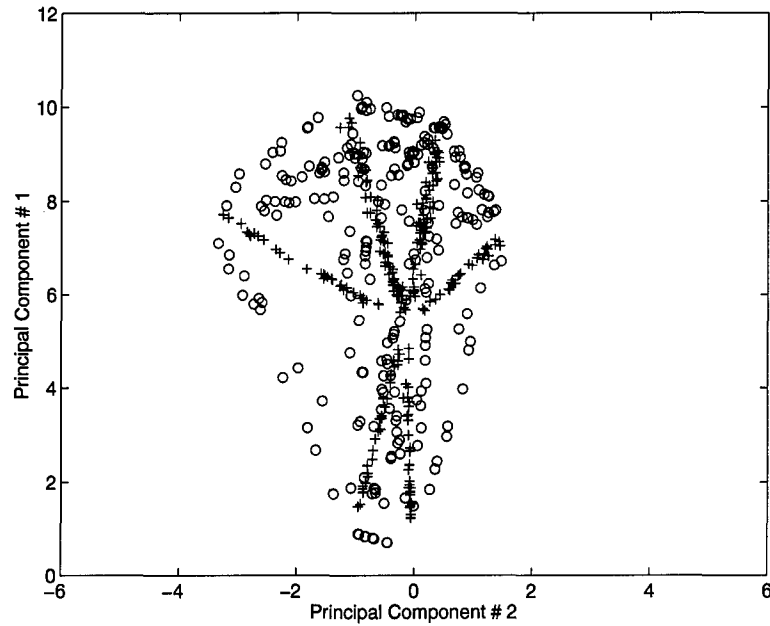


Figure 29. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 3 data shown with pluses and other material data shown with circles.

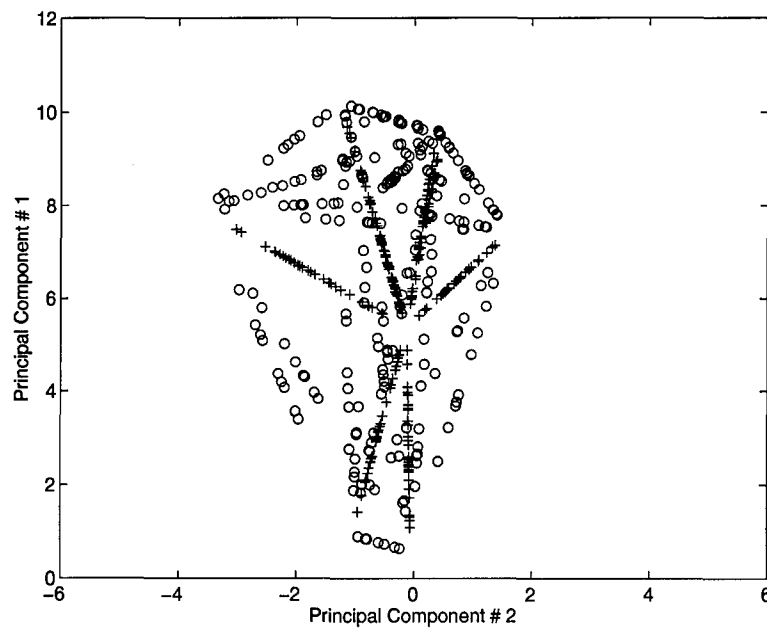


Figure 30. Two Material Training Data Projected into PC Plane: High SNR Case. Material 3 data shown with pluses and other material data shown with circles.

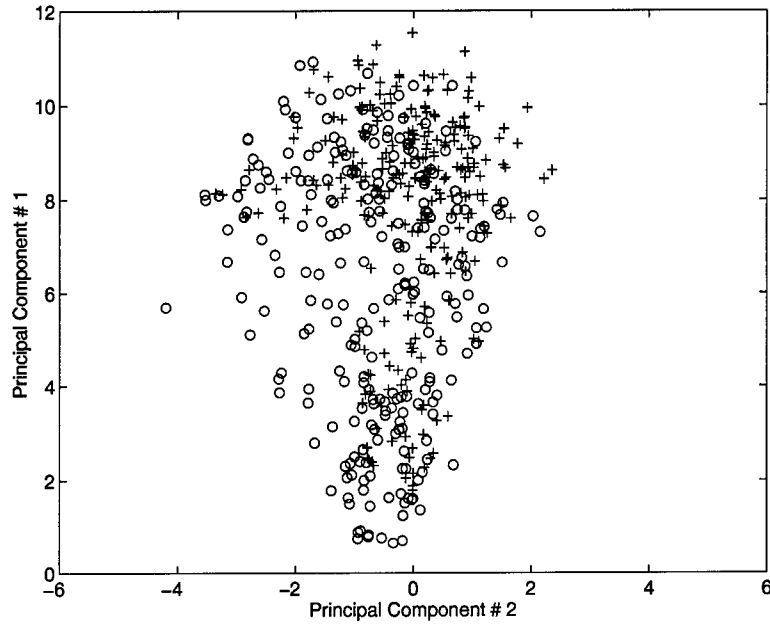


Figure 31. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 4 data shown with pluses and other material data shown with circles.

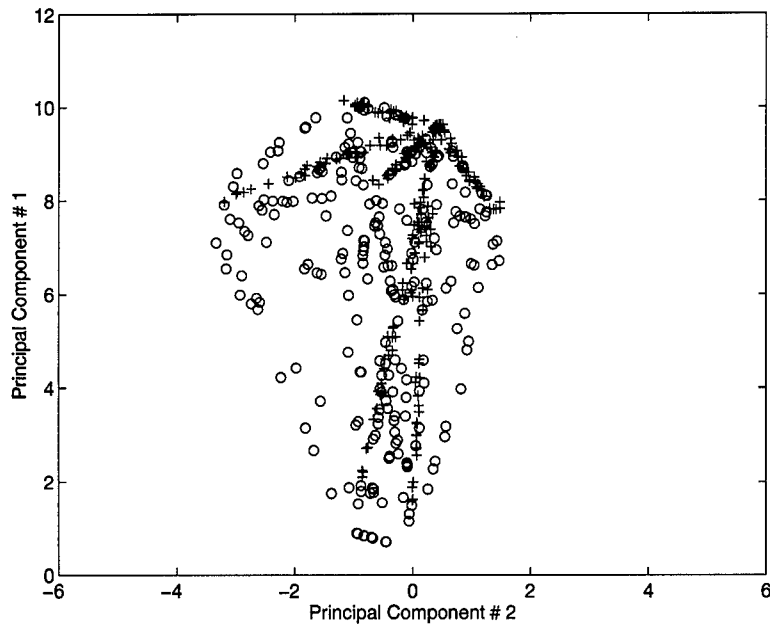


Figure 32. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 4 data shown with pluses and other material data shown with circles.

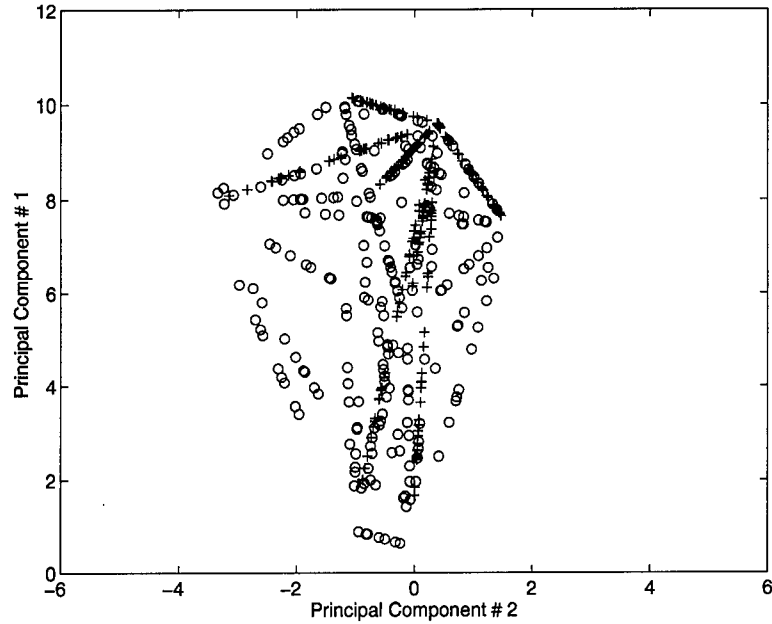


Figure 33. Two Material Training Data Projected into PC Plane: High SNR Case. Material 4 data shown with pluses and other material data shown with circles.

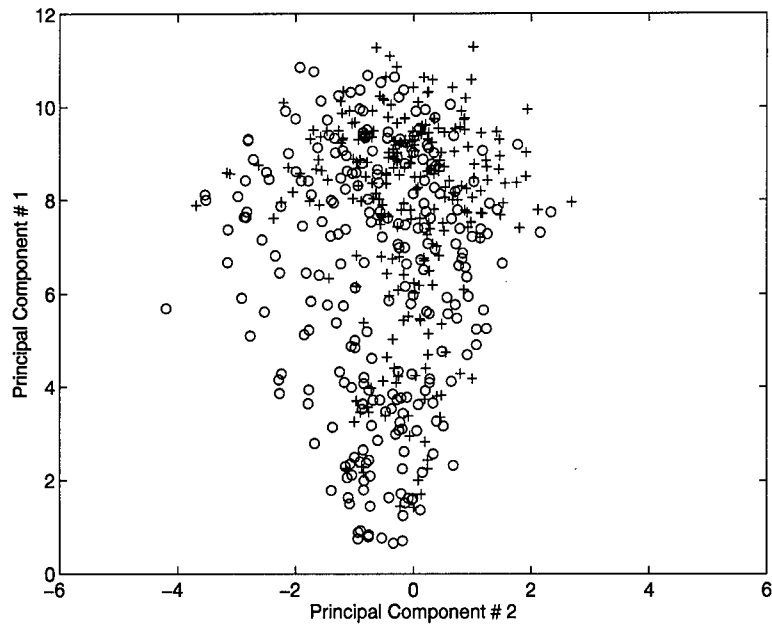


Figure 34. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 5 data shown with pluses and other material data shown with circles.

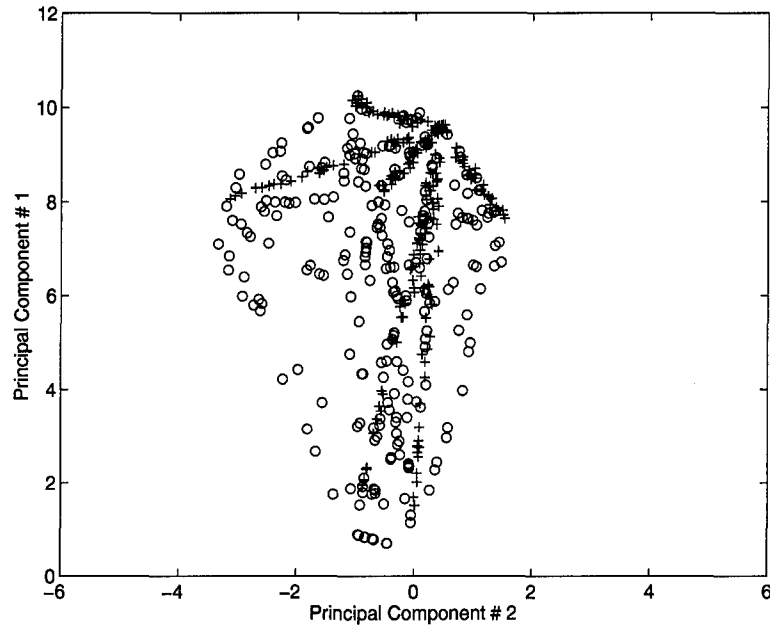


Figure 35. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 5 data shown with pluses and other material data shown with circles.

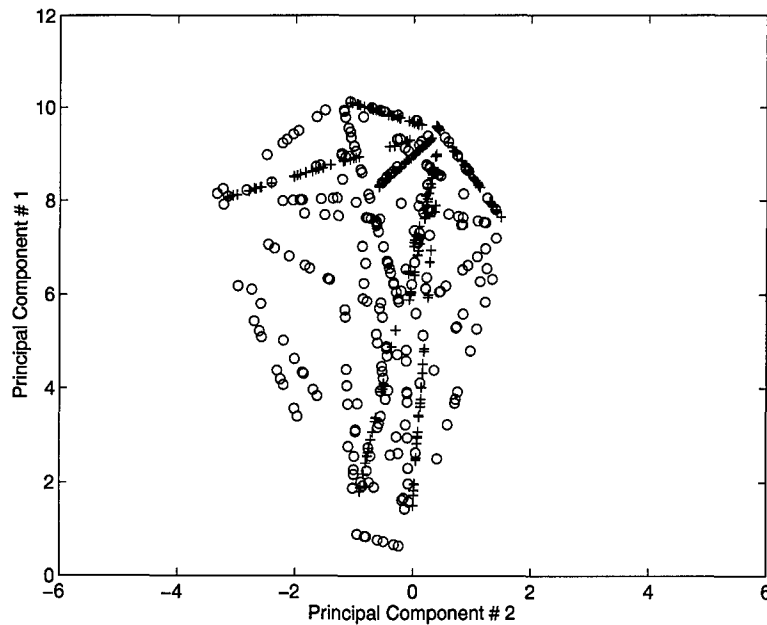


Figure 36. Two Material Training Data Projected into PC Plane: High SNR Case. Material 5 data shown with pluses and other material data shown with circles.

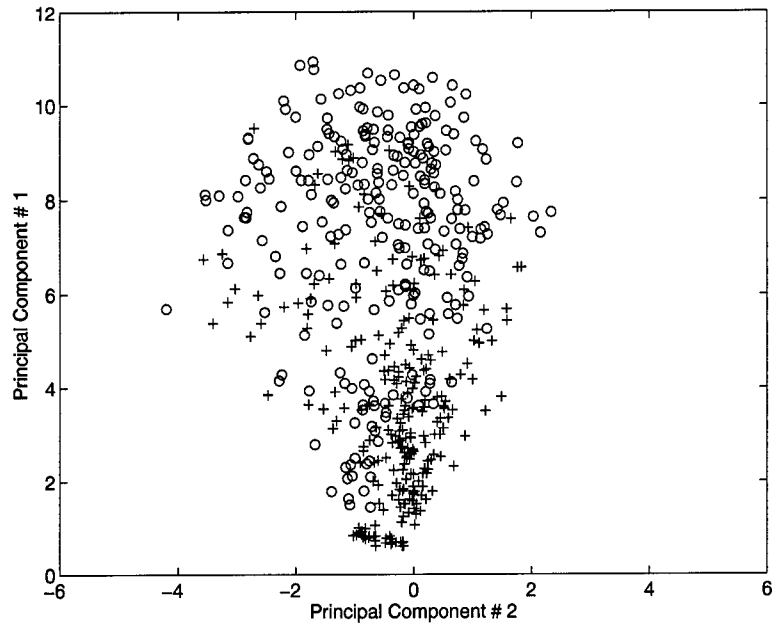


Figure 37. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 6 data shown with pluses and other material data shown with circles.

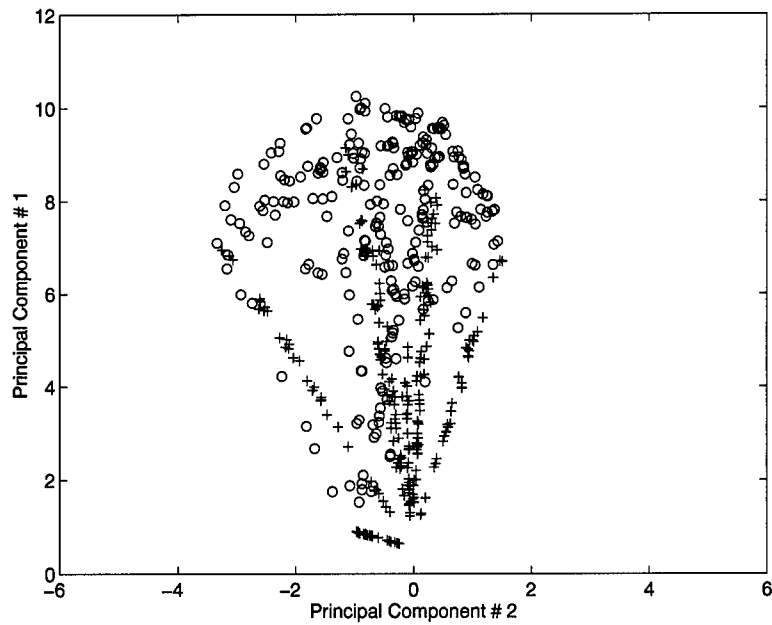


Figure 38. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 6 data shown with pluses and other material data shown with circles.

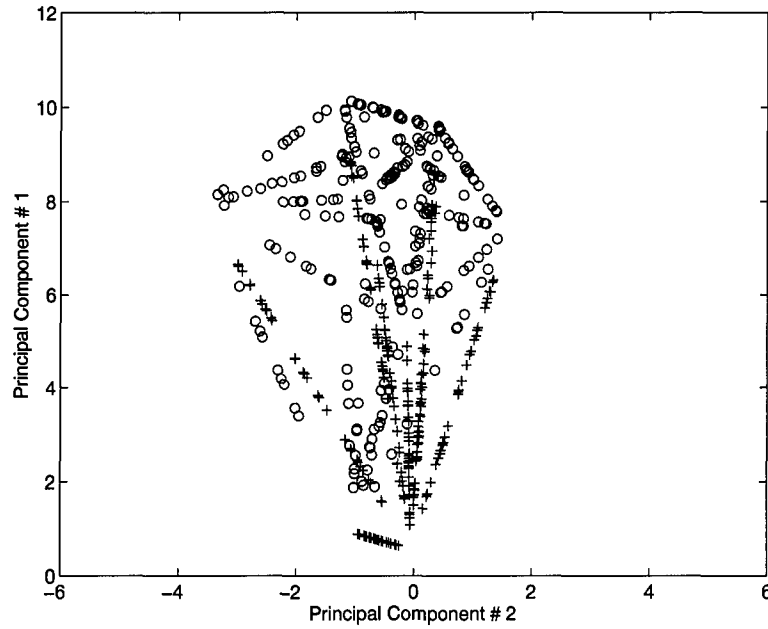


Figure 39. Two Material Training Data Projected into PC Plane: High SNR Case. Material 6 data shown with pluses and other material data shown with circles.

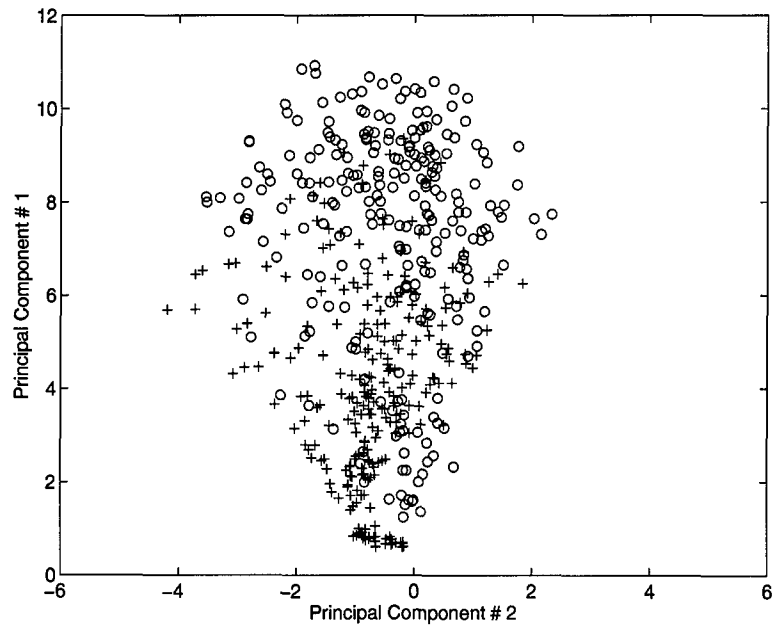


Figure 40. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 7 data shown with pluses and other material data shown with circles.

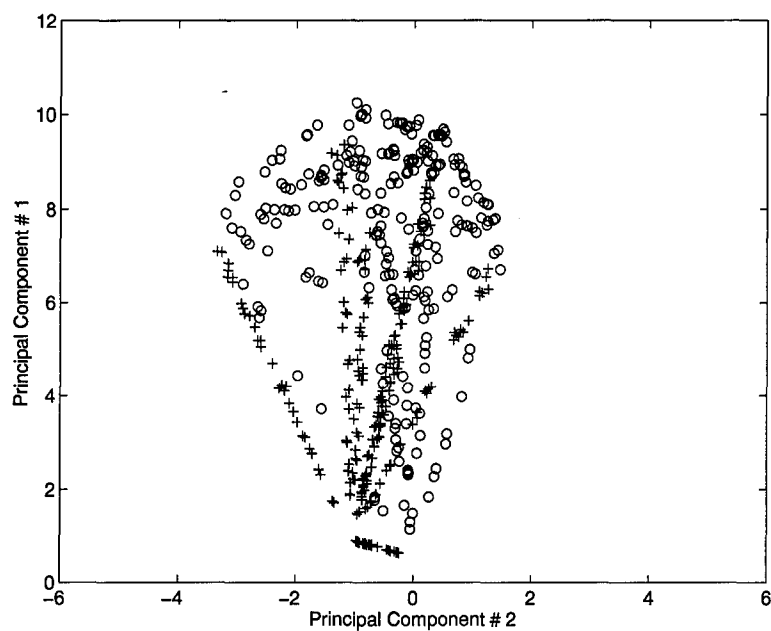


Figure 41. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 7 data shown with pluses and other material data shown with circles.

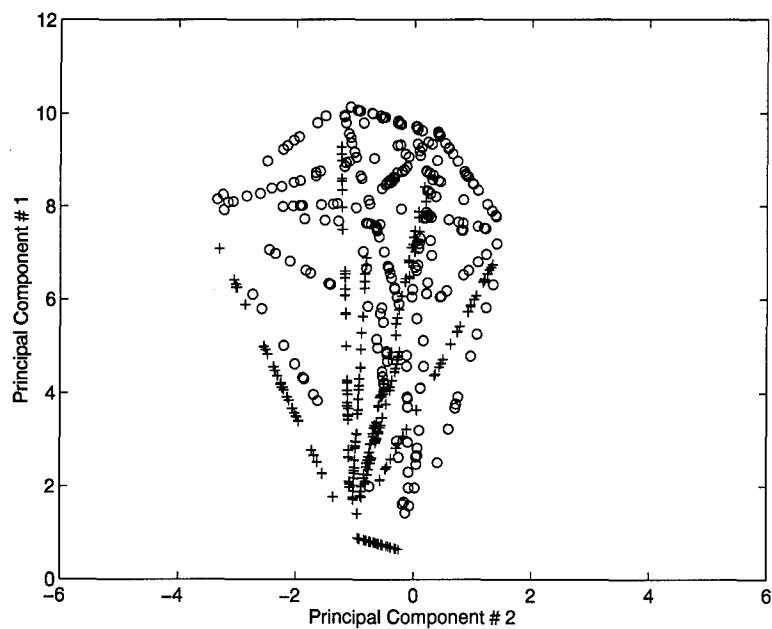


Figure 42. Two Material Training Data Projected into PC Plane: High SNR Case. Material 7 data shown with pluses and other material data shown with circles.

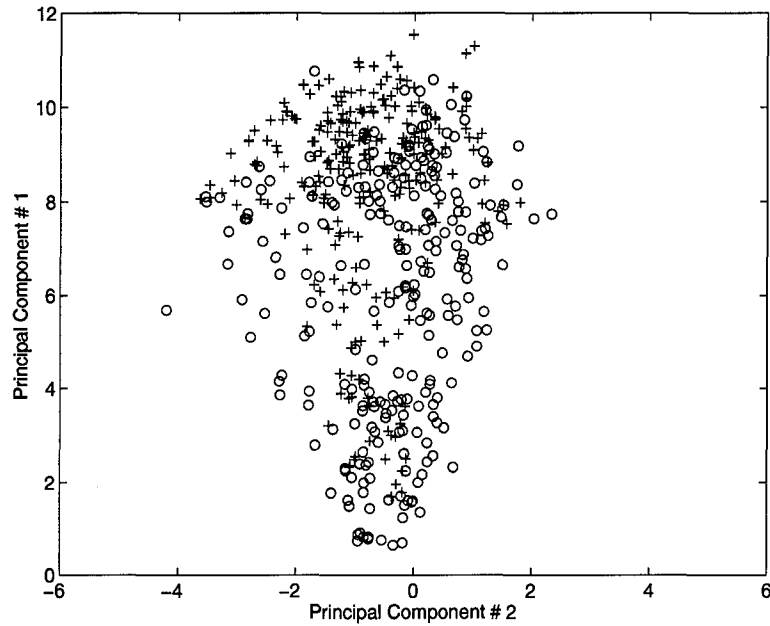


Figure 43. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 8 data shown with pluses and other material data shown with circles.

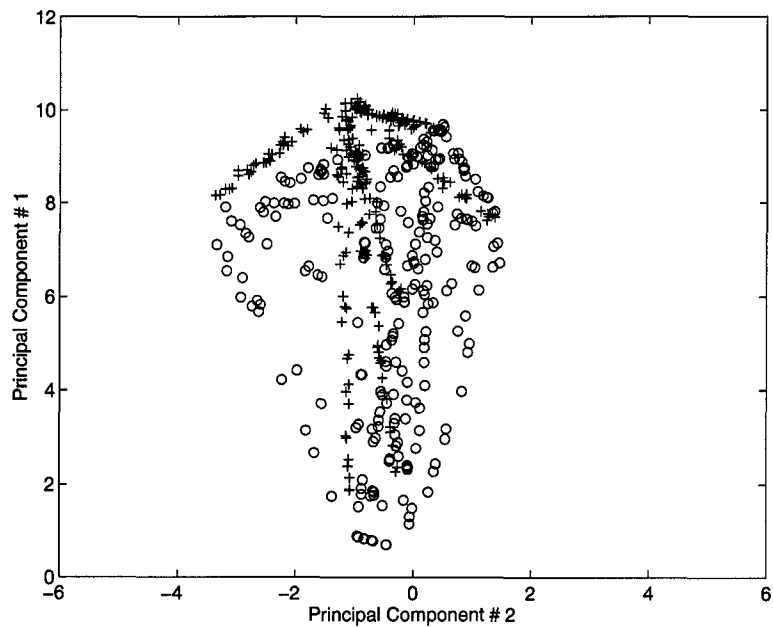


Figure 44. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 8 data shown with pluses and other material data shown with circles.

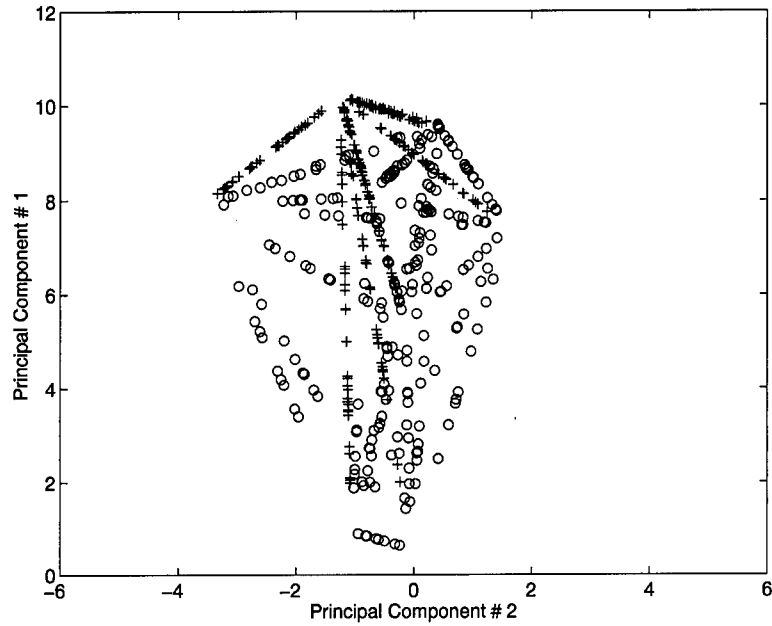


Figure 45. Two Material Training Data Projected into PC Plane: High SNR Case. Material 8 data shown with pluses and other material data shown with circles.

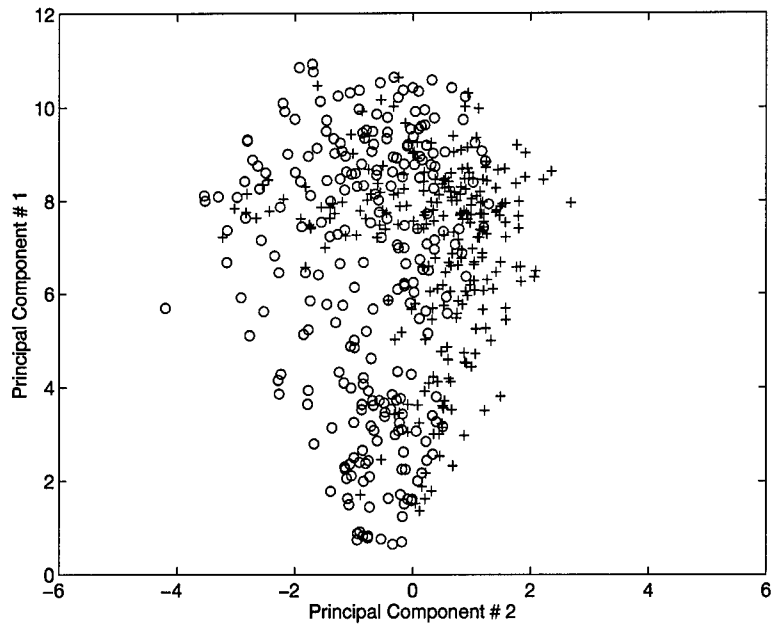


Figure 46. Two Material Training Data Projected into PC Plane: Low SNR Case. Material 9 data shown with pluses and other material data shown with circles.

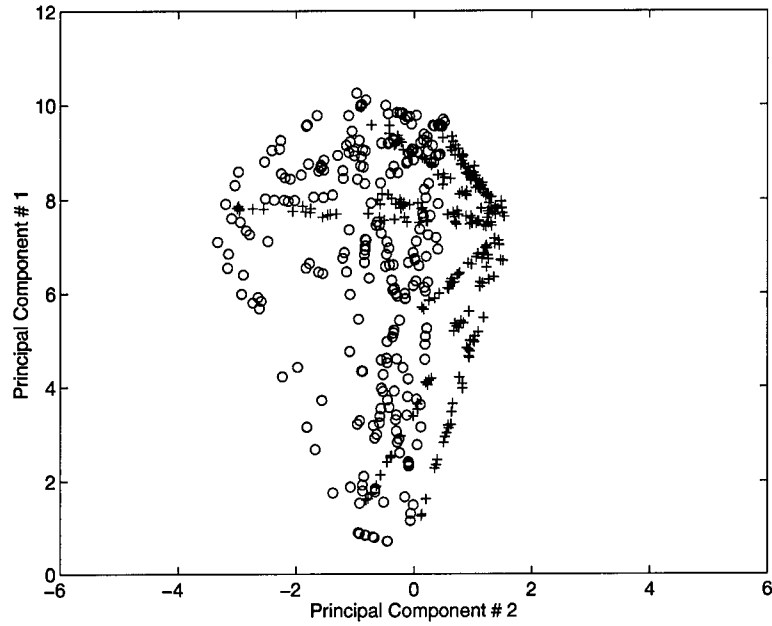


Figure 47. Two Material Training Data Projected into PC Plane: Mid SNR Case. Material 9 data shown with pluses and other material data shown with circles.

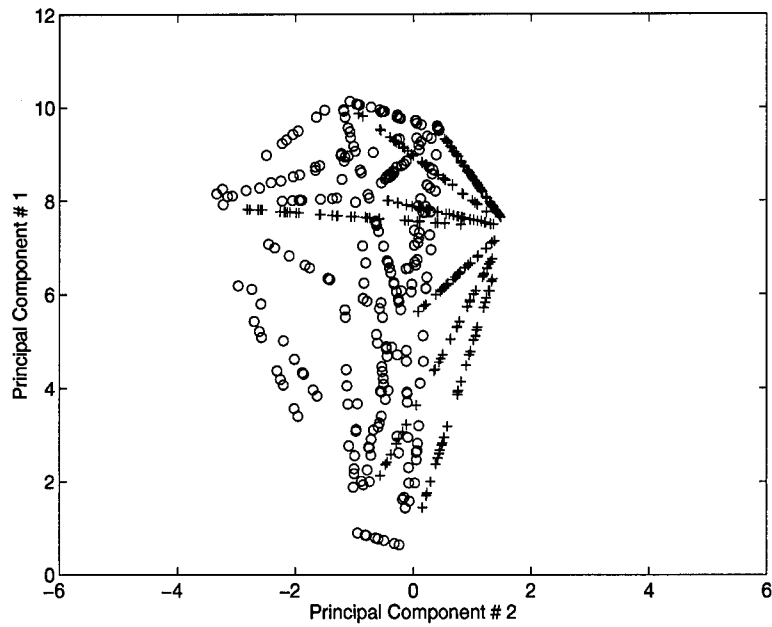


Figure 48. Two Material Training Data Projected into PC Plane: High SNR Case. Material 9 data shown with pluses and other material data shown with circles.

A.3 ANN Rankings for Two Materials whose Spectra are in the Composite Spectrum

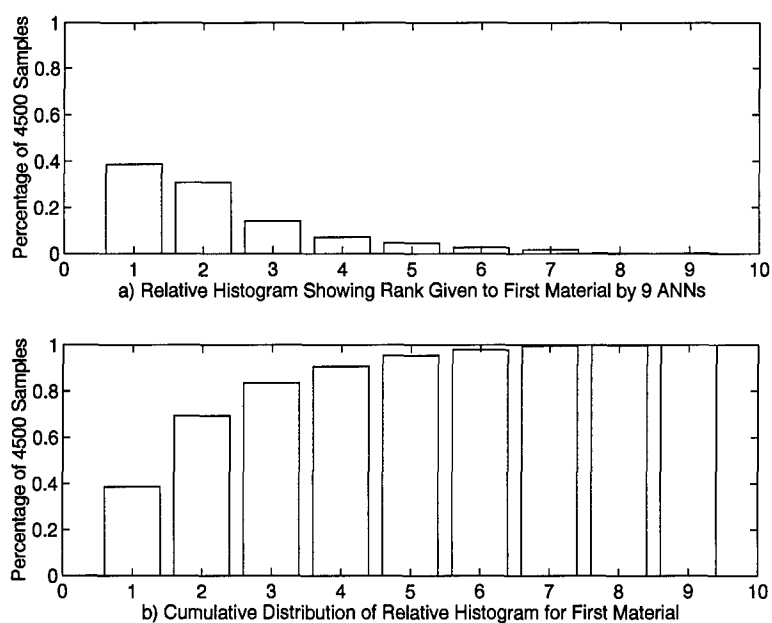


Figure 49. ANN Rankings for First Material in Two Material Composite for the Mid SNR Case.

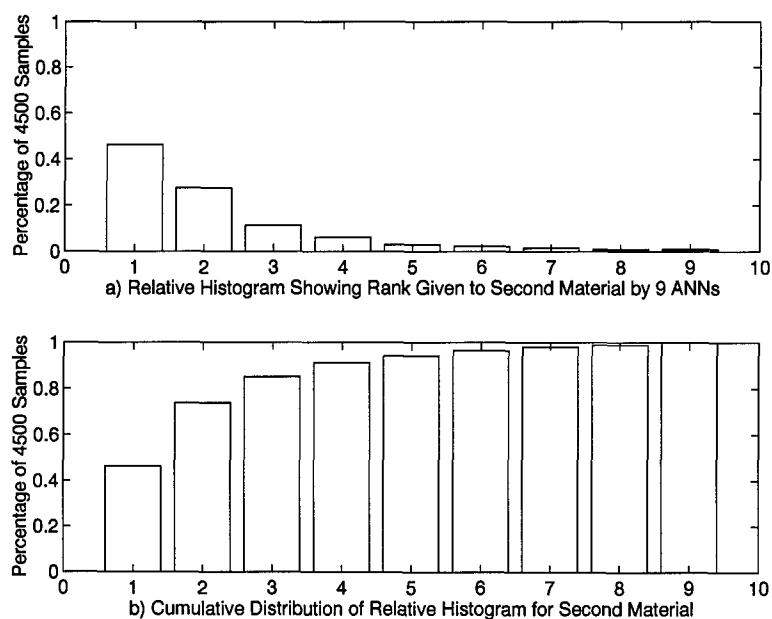


Figure 50. ANN Rankings for Second Material in Two Material Composite for the Mid SNR Case.

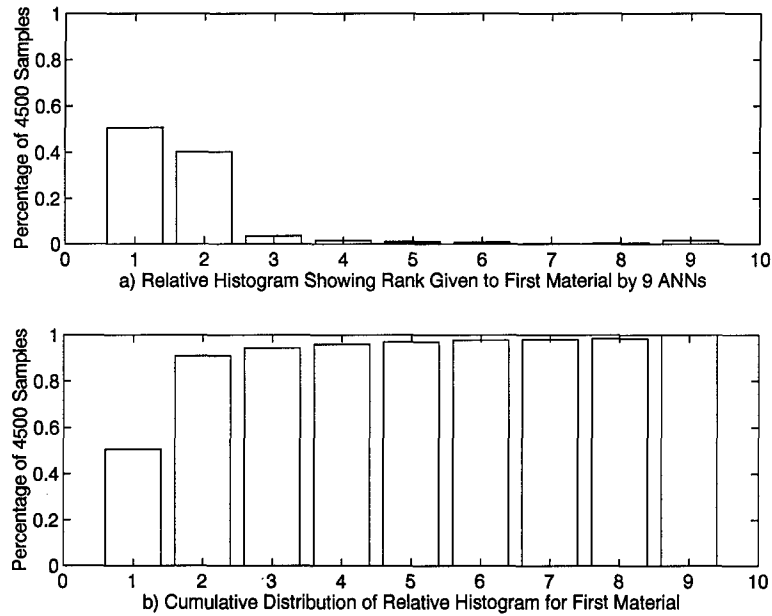


Figure 51. ANN Rankings for First Material in Two Material Composite for the High SNR Case.

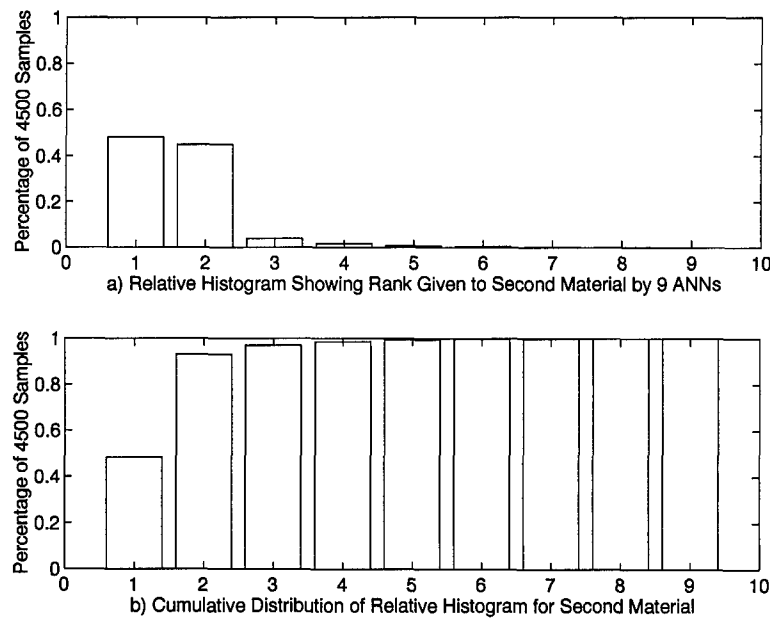


Figure 52. ANN Rankings for Second Material in Two Material Composite for the High SNR Case.

A.4 Errors in Estimation of Percent Composition: Hybrid Approach

Table 17. Errors in Estimates using Hybrid Approach; Mid SNR Case. n spectra were used in the database entered into the linear estimator. ϵ_1 refers to the RMS error of the estimate when spectrum of material present in composite. ϵ_0 refers to RMS error of estimate when spectrum of material not present in composite. ϵ_t is the total RMS error for the estimate. The mean values are the average RMS error across all nine materials.

Material	n=3			n=4			n=5		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
1	.0985	.0239	.0510	.0819	.0095	.0395	.0516	.0069	.0251
2	.2420	.1539	.1773	.2255	.1147	.1467	.2089	.0917	.1274
3	.2550	.1851	.2027	.2155	.1253	.1501	.1817	.0801	.1110
4	.1731	.0927	.1155	.1181	.0663	.0807	.0744	.0492	.0558
5	.0734	.0207	.0391	.0673	.0169	.0351	.0644	.0124	.0323
6	.2385	.0747	.1303	.1764	.0555	.0965	.0878	.0525	.0621
7	.1690	.0734	.1027	.1190	.0465	.0695	.1011	.0299	.0545
8	.2201	.0783	.1247	.1690	.0816	.1074	.1400	.0769	.0946
9	.1995	.0597	.1078	.1090	.0240	.0556	.0827	.0152	.0412
<i>mean</i>	.1855	.0847	.1168	.1424	.0600	.0868	.1103	.0461	.0671
Material	n=6			n=7			n=8		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
1	.0325	.0073	.0166	.0188	.0092	.0120	.0145	.0108	.0117
2	.2011	.0742	.1152	.1936	.0653	.1079	.1863	.0805	.1129
3	.2003	.0351	.0994	.2391	.0262	.1151	.2569	.0177	.1221
4	.0538	.0403	.0437	.0519	.0385	.0419	.0534	.0362	.0407
5	.0607	.0101	.0300	.0577	.0133	.0296	.0492	.0139	.0263
6	.0595	.0646	.0635	.0455	.0686	.0642	.0437	.0630	.0593
7	.0935	.0222	.0482	.0846	.0253	.0457	.0699	.0260	.0401
8	.1152	.0758	.0861	.1055	.0712	.0801	.1165	.0672	.0808
9	.0530	.0150	.0283	.0371	.0169	.0230	.0284	.0186	.0212
<i>mean</i>	.0966	.0383	.0590	.0926	.0372	.0577	.0910	.0371	.0572

Table 18. Errors in Estimates using Hybrid Approach; High SNR Case. n spectra were used in the database entered into the linear estimator. ϵ_1 refers to the RMS error of the estimate when spectrum of material present in composite. ϵ_0 refers to RMS error of estimate when spectrum of material not present in composite. ϵ_t is the total RMS error for the estimate. The mean values are the average RMS error across all nine materials.

Material	n=3			n=4			n=5		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
1	.0137	.0013	.0066	.0110	.0006	.0052	.0063	.0004	.0030
2	.1548	.0774	.0999	.1443	.0957	.1084	.1229	.0967	.1031
3	.1768	.1212	.1355	.1015	.0974	.0983	.0582	.0456	.0487
4	.0507	.0082	.0250	.0370	.0024	.0176	.0268	.0031	.0129
5	.0226	.0060	.0119	.0030	.0042	.0040	.0024	.0072	.0065
6	.0861	.0181	.0436	.0642	.0104	.0316	.0432	.0078	.0215
7	.0505	.0210	.0302	.0256	.0089	.0144	.0064	.0125	.0114
8	.1288	.0219	.0637	.1257	.0276	.0641	.1184	.0417	.0668
9	.0212	.0095	.0131	.0069	.0038	.0047	.0020	.0043	.0039
mean	.0784	.0316	.0477	.0577	.0279	.0387	.0429	.0244	.0309
Material	n=6			n=7			n=8		
	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t	ϵ_1	ϵ_0	ϵ_t
1	.0037	.0008	.0019	.0014	.0036	.0032	.0016	.0038	.0034
2	.0777	.0962	.0924	.0170	.0913	.0809	.0214	.0817	.0728
3	.0677	.0115	.0335	.0692	.0073	.0333	.0771	.0024	.0364
4	.0129	.0039	.0070	.0115	.0046	.0068	.0060	.0061	.0061
5	.0026	.0037	.0035	.0027	.0029	.0029	.0028	.0021	.0023
6	.0175	.0059	.0098	.0132	.0031	.0068	.0056	.0067	.0065
7	.0042	.0059	.0056	.0041	.0052	.0050	.0032	.0037	.0036
8	.1164	.0295	.0607	.1126	.0049	.0533	.1014	.0073	.0482
9	.0020	.0025	.0024	.0021	.0032	.0030	.0021	.0027	.0026
mean	.0339	.0178	.0241	.0260	.0140	.0217	.0246	.0129	.0202

Bibliography

1. Baum, Eric B. "What Size Net Gives Valid Generalization," *Neural Computation*, 1 (1989).
2. Bell, Robert J. *Introductory Fourier Transform Spectroscopy*. New York: Academic Press, 1972.
3. Boardman, Joe W. "Inversion of Imaging Spectrometry Data using Singular Value Decomposition." *IGARSS '89, 12th Canadian Symposium on Remote Sensing* 4. 2069-2072. 1989.
4. Boyd, Robert W. *Radiometry and the Detection of Optical Radiation*. New York: John Wiley and Sons, 1983.
5. Cybenko, G. "Approximation by Superpositions of a Sigmoidal Function," *Mathematics of Control, Signals, and Systems*, 2(4):303-314 (1989).
6. D.N.Chen, K. Motoshima, M.M.Downs E.Desurvire. "Reentrant Sagnac Fiber Gyroscope with a Recirculating Delay Line using an Erbium-Doped Fiber Amplifier," *IEEE Photonics Technology Letters*, 4(7):813-817 (July 1992).
7. Duda, Richard O. and Peter E. Hart. *Pattern Classification and Scene Analysis*. New York: John Wiley and Sons, 1973.
8. E.E.Bell. "Measurement of the Far Infrared Optical Properties of Solids with a Michelson Interferometer Used in the Asymmetric Mode: Part I, Mathematical Formulation," *Infrared Physics*, 6:57-74 (1966).
9. Faires, Lynda M. "Fourier Transforms for Analytical Atomic Spectroscopy," *Analytical Chemistry*, 58(9):1023A-1034A (August 1986).
10. F.A.Kruse, A.B.Lefkoff, J.B.Dietz. "Expert System-Based Mineral Mapping in Northern Death Valley, California/Nevada, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sensing of the Environment*, 44:309-336 (1993).
11. Foley, Donald H. "Consideration of Sample and Feature Size," *IEEE Transactions on Information Theory*, IT-18(5):618-626 (September 1972).
12. Fukunaga, Keinosuke. *Introduction to Statistical Pattern Recognition*. Boston: Academic Press, 1990.
13. Goodman, Joseph W. *Statistical Optics*. New York: John Wiley and Sons, 1985.
14. Hrovat, Daniel. *Hyperspectral Analysis of Space Objects: Signal to Noise Evaluation*. MS thesis, Air Force Institute of Technology, December 1993.
15. Hush, Don R. and Bill G. Horne. "Progress in Supervised Neural Networks," *IEEE Signal Processing Magazine* (January 1993).
16. Inc., Math Works. *The Student Edition of MATLAB, For MS-DOS Personal Computers*. Englewood Cliffs, New Jersey: Prentice Hall, 1992.

17. Jain, Anil K. and Jianchang Mao. "Neural Networks and Pattern Recognition,"
18. Johnson, Richard A. and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Englewood Cliffs, New Jersey: Prentice Hall, 1982.
19. Loewenstein, Ernest V. "Fourier Spectroscopy: An Introduction," *Aspen International Conference on Fourier Spectroscopy*, 3-17 (1970). AFCRL-71-0019, 5 Jan 71, Special Report Number 114.
20. Martin, Curtis E. *Non-Parametric Bayes Error Estimation for UHRR Target Identification*. MS thesis, Air Force Institute of Technology, Wright-Patterson AFB, December 1993.
21. Michael E. Forman, W.H.Steel and George A. Vanasse. "Correction of Asymmetric Interferograms Obtained in Fourier Spectroscopy," *Journal of the Optical Society of America*, 56(1):59-63 (January 1966).
22. Parsons, Thomas W. *Voice and Speech Processing*. New York: McGraw-Hill, 1987.
23. Parzen, Emanuel. "On Estimation of a Probability Density Function and Mode," *Ann. Math. Stat.*, 33 (1962).
24. R.A.Williams and W.S.Chang. "Resolution and Noise in Fourier-Transform Spectroscopy," *Journal of the Optical Society of America*, 56(2):167-170 (February 1966).
25. Rogers, Steven K. "An Introduction to Biological and Artificial Neural Networks." Wright-Patterson AFB, Air Force Institute of Technology, 1990.
26. Rosenblatt. *Principles of Neurodynamics*. Spartan Books, 1959.
27. Sakai, Hajime. "Consideration of the Signal-to-Noise Ratio in Fourier Spectroscopy," *Aspen International Conference on Fourier Spectroscopy*, 19-41 (1970). AFCRL-71-0019, 5 Jan 71, Special Report Number 114.
28. Schalkoff, Robert. *Pattern Recognition: Statistical, Structural, and Neural Approaches*. New York: John Wiley and Sons, 1992.
29. S.C.Huang and Y.F.Huang. "Bounds on the Number of Hidden Neurons in Multilayer Perceptrons," *IEEE Transactions on Neural Networks*, 2(1):47-55 (1991).
30. Takayuki Okamoto, Satoshi Kawata and Shigeo Minami. "Fourier Transform Spectrometer with a Self-Scanning Photodiode Array," *Applied Optics*, 23(2):269-273 (January 1984).
31. Takayuki Okamoto, Satoshi Kawata and Shigeo Minami. "Optical Method for Resolution Enhancement in Photodiode Array Fourier Transform Spectroscopy," *Applied Optics*, 24(23):4221-4225 (December 1985).
32. W.H.Steel. *Interferometry* (Second Edition). Cambridge: Cambridge University Press, 1983.
33. Yoshihara, Kunio and Atsuo Kitade. "Holographic Spectra Using a Triangle Path Interferometer," *Japanese Journal of Applied Physics*, (6):116 (1967).

Vita

Captain Eugene L. Caudill was born on 16 July 1963 in Santa Cruz, California. He graduated from high school in Dallas, Texas in 1981 and enlisted in the USAF in March 1982. Upon completion of technical training, he was assigned to the 46th Communications Group, Barksdale AFB, Louisiana as a Ground Radio Equipment Repair Specialist. After completing over 30 collegiate credit hours, he applied to the Air Force Airman Education and Commissioning Program (AECP). He completed his undergraduate requirements at Texas A & M University and received the degree of Bachelor of Science in Electrical Engineering in December 1989. Upon completion of Officer Training School, he received a commission in the USAF in April 1990. From May 1990 to May 1993 he was assigned to the Phillips Laboratory, Advanced Imaging Technology Branch, Kirtland AFB, New Mexico. In June 1993, Captain Caudill entered AFIT to begin work toward a Master of Science degree in Electrical Engineering.

Permanent address: 3509 Woodoak
Balch Springs, Texas 75180

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE December 1994		3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE Satellite Surface Material Composition from Synthetic Spectra				5. FUNDING NUMBERS	
6. AUTHOR(S) Eugene L. Caudill, Capt, USAF					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology, WPAFB OH 45433-6583				8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/GEO/ENP/94D-02	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Maj Susan Durham PL/LIMI, KAFB, NM 71117-6008				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The objective of this research was to determine if measurements from a Sagnac interferometer could provide reliable estimates of satellite material composition. The Sagnac interferometer yields a spatial interferogram that can be sampled by a linear detector array. The interferogram is related to the spectrum of the source through a Fourier transform. Here, spectral reflectivities of nine common satellite materials were used to simulate the spectrum one obtains from an ideal Sagnac interferometer in the beam-train of a ground-based telescope whose mission is to view satellites. The signal-to-noise ratio of the spectrum was varied to simulate the effect of range variation between the sensor and the satellite. The simulated spectra consisted of a linear mixture of spectra from two of the nine materials. Three different architectures were developed and their performances compared. One of the three architectures consisted of nine artificial neural networks (ANNs), one for each material, and a linear estimator that estimated the satellite surface area attributable to each material. This method estimates the material composition by using a classifier to identify the materials contributing to the mixture, then eliminating unlikely contributors to the mixture before performing a constrained linear estimate. It is shown that due to high classification errors, the system using solely a linear estimator provides the estimate with the lowest errors.					
14. SUBJECT TERMS Sagnac, Multi-Layer Perceptron, Abundance, Principal Components				15. NUMBER OF PAGES 117	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL		